The impact of knowledge management processes on organizational resilience: data mining as an instrument of measurement.

FRELAS, M.

2017

THE IMPACT OF KNOWLEDGE MANAGEMENT PROCESSES ON ORGANIZATIONAL RESILIENCE: DATA MINING AS AN INSTRUMENT OF MEASUREMENT

MICHAEL FRELAS

A thesis submitted in partial fulfillment of the
requirements of
The Robert Gordon University
for the degree of Doctor of Information Science

April 2017

ABSTRACT

The aim of the research conducted for this thesis is to test the feasibility of using data mining (DM) to assess the relationship between and the impact of knowledge management (KM) on organizational resilience (OR).

The emphasis currently placed on the value of intangible assets by private sector organizations and the recent increase in the use of data mining technologies are the key drivers in this evaluation of the use of data mining tools as an alternative to classical statistics when measuring intangibles.

Data was collected using a questionnaire that was sent to the senior executives of a number of mid-sized companies located in the mid-west of the USA. Using Microsoft's SQL Server's Analytical Services (MSSAS) and the data provided by the respondents, five predictive models are built to test the suitability of the MSSAS' DM tool for assessing the relationships between and the impact of KM on OR.

Of the five models constructed as part of this research, four classification models (two Naïve Bayes models, one neural network model, and one decision tree model) and one clustering model were found to be suitable tools for capturing the intricate relationships that exist between KM and OR. These models made it possible to evaluate the strengths of the relationships between KM and OR and to identify which KM processes contribute, and to what extent, to OR. In addition, the models enabled the collation of predicted OR scores, based on the responses given in the questionnaire. Finally, this research identifies some of the key challenges associated with using DM as a measurement instrument for assessing the relationship between and the impact of KM on OR.

This research makes a number of significant contributions to the existing body of knowledge. It contributes to the understanding of the impact of KM on OR, to the understanding of the methods used to measure such impact and to the processes involved in measuring such impact using DM. From a practitioner perspective, this research contributes to the understanding of OR and provides a framework for achieving OR within an organizational context.

Organizational Resilience, Knowledge Management, Data Mining, Performance

ACKNOWLEDGMENTS

Some people may say this project lasted as long as a good marriage does nowadays, as it has been exactly seven years since the commencement of this project.

As with most marriages, there were a number of ups and downs, but nothing that one would not expect to find in the books on 'how to get a PhD.' It seems as if I have encountered every stage discussed in such publications: enthusiasm, isolation, interest in the work, boredom, and frustration. (Then, repeat.)

I have to express my special gratitude to Professor Simon Burnett who, as a primary supervisor, has truly been a great help in assisting me to get through each stage of the doctoral program and encouraging me to stay in the program at some critical moments. I also appreciate Professor Burnett's willingness to discuss my research at various times of day and in various parts of the world.

This work would not be near the quality it is now had it not been for Professor Dorothy Williams and the time she spent away from her retirement reading it and providing invaluable comments.

Special thanks also go to Dr. Iain Pirie for being a sounding board regarding the data modeling questions and to all of the other great people, both from RGU and beyond, who have in some way supported me on the way to the completion of this research.

The work would have taken much longer to complete if were not for the flexible/remote work schedule provided by my current employer Integrity Payment Systems; thank you for accommodating my needs.

Finally, I must express my appreciation to my kids, Frank and Anthony, for understanding Dad's need to work on "academic stuff" and for keeping the noise levels down. Also, thank you Aleks.

In the end, the completion of this work would not have been possible were it not for my beautiful wife, Ewa. Thank you, Ewa, for managing expectations and time at home, taking on a larger role within the family, putting your career on hold and keeping our marriage alive and well during the last seven years.

CONTENTS

TABLES

FIGURES

APPENDICES

ABBREVIATIONS

7S	- Seven Groups of Questionnaire Questions
ANOVA	- Analysis of Variance
AMC	- Awareness-Motivation-Capability
API	- Application Programming Interface
AQ	- Access Quality
AR	- Applied Research
AVE	- Average Variance Extracted
BI	- Business Intelligence
BI&A	- Business Intelligence and Analytics
BIS	- Business Intelligence System
BN	- Bayesian Network
BNC	- Bayesian Network Classifier
BR	- Basic Research
CA	- Competitive Advantage
CI	- Covering Index
CO	- Connect Competence
CR	- Creative Competence
CRM	- Customer Resource Management
DDDM	- Domain-Driven Data Mining
DDDM-PD	- Domain-Driven Data Mining Pattern Discovery
DDID-PD	- Domain-Driven In-Depth Pattern Discovery [Subset of DDDM-PD]

DE	- Decide Competence
DIKAR	- Data, Information, Knowledge, Action, Results
DIKW	- Data, Information, Knowledge, Wisdom
DM	- Data Mining
DMX	- Data Mining Extension
DSS	- Decision Support System
DT	- Decision Trees
DW	- Data Warehouse
EDM	- Extension Data Mining
EIS	- Executive Information System
EM	- Expectation Maximization
ER	- Enterprise Resilience
ETL	- Extract, Transpose, Load
EX	- Exploit Competence
FIVA	- Framework of Intangible Valuation Areas
FKMS	- Financial Knowledge Management System
FNN	- Fuzzy Neural Network
GA	- Genetic Algorithm
GDP	- Gross Domestic Product
IC	- Intellectual Capital
ICT	- Information Communication Technology
IM	- Information Management
IQ	- Information Quality

IQR	- Interquartile Range
IS	- Information Science
IT	- Information Technology
KAVS	- Knowledge Asset Value Spiral
KBV	- Knowledge-Based View
KDD	- Knowledge Discovery in Databases
KLC	- Knowledge Life Cycle
KM	- Knowledge Management
KMS	- Knowledge Management System
KMP	- Knowledge Management Processes
KVS	- Knowledge Valuation System
KW	- Knowledge Warehouse
LE	- Learn Competence
LI	- Link Competence
LVQ	- Learning Vector Quantization
MADM	- Multiple Attribute Decision-Making
MBV	- Market-Based View
MCDM	- Multiple Criteria Decision-Making
MIS	- Management Information System
MS	- Microsoft
MSSAS	- Microsoft SQL Server Analytical Services
NN	- Neural Network
OA	- Operational Agility

OL	- Organizational Learning
OLAP	- On-Line Analytical Processing
OpR	- Operation Research
OP	- Organizational Performance
OR	- Organizational Resilience
PA	- Predictive Analytics
PE	- Performance Competence
RAKID	- Results, Action, Knowledge, Information, Data
RBT	- Resource-Based Theory
RBV	- Resource-Based View
RM	- Revenue Management
ROA	- Return on Assets
RQ	- Research Question
RST	- Rough Set Theory
SBV	- Stakeholder Based View
SCRes	- Supply Chain Resilience
SKM	- Strategic Knowledge Management
SMB	- Small/Medium (sized) Business
SOFM	- Self-Organizing Feature Map
SQL	- Structured Query Language
SR	- Systematic Resilience
SSIS	- SQL Server Integration Services (SQL Server is a product of Microsoft Corporation

CHAPTER ONE: INTRODUCTION

1.1 Research and the Research Problem

This thesis focuses on how data mining can assist in identifying the intricate relationships that exist between knowledge management and organizational resilience and on measuring the impact of knowledge management on organizational resilience. The publications of Davenport (Davenport & Harris, 2007) and Davenport et al. (2010), which discuss a firm's ability to compete based on analytics, had the greatest impact on the origins of this research. These publications prompted the following questions: Is analytics the factor that allows for some organizations to always outperform their competition, regardless of business conditions? Does analytics contribute to such organizational resilience? Can analytics identify the key factors or processes that must be present in organizations that constantly perform well (in other words, those that are resilient)?

Given the great ability of modern companies to emulate their competitor's resources, there must be another factor, besides tangible resources, that accounts for their success. The missing component that may be considered as contributing to organizational resilience was identified as a result of realizing the power of knowledge and the need for knowledge management discussed in the work of King (2009), who stated that 'just as human beings are unable to draw on the full potential of their brains, organizations are generally not able to fully utilize the knowledge that they possess.' Thus, if utilization of knowledge is important in order for an organization to be resilient, then, as is usual in the business world, it is also necessary to measure the impact of such utilization on a business. The utilization of knowledge and the measurement of the impact of utilized knowledge on an organization fall under the domain of the knowledge management processes; therefore, this work focuses on knowledge management (KM) processes rather than on knowledge itself, which, when not applied, most likely provides minimal benefits to an organization.

With the resilience of an organization being key to business success and the KM processes being the key intangible resilience success factor, this thesis addresses the need for and develops a methodological approach for achieving such resilience.

While the KM processes and the business success factors have been extensively researched, the impact of the KM processes on an organization's resilience is not fully understood. Thus, this thesis sets out to identify the methods by which such impact can be measured and the relationships between KM and organizational resilience can be detected, with the purpose of allowing firms to use this information to improve their resilience.

Due to the increasing significance of data mining (DM) techniques that offer the derivation of valuable business insights, among other things, the realization came that these techniques could be utilized to identify the relationships between and impact of KM on organizational resilience, as these are difficult to detect with traditional statistical tools.

The realization of the importance of organizational resilience and its key intangible KM component, viewed through the DM lens, became the focus of this research. The aim of this thesis is thus to test the feasibility of using data mining to assess the relationship between and impact of knowledge management on organizational resilience (as an ability to always perform well, regardless of the business environment). Given the ease by which tangible resources can be acquired, the KM processes became the focus for achieving such resilience.

The research aim, objectives and methodology of this thesis are further discussed in Chapter 4 and are also briefly explored in the following sections of this chapter.

1.2 Aim and Research Context

As mentioned in the previous section, the main aim of the applied, multidisciplinary research conducted for the purpose of this thesis is to determine what insights analytics can provide with respect to the impact of the KM processes on organizational well-being (the definition of organizational resilience, as well as other key definitions for this work, is provided in Chapter 2).

Specifically, the aim of this applied, multidisciplinary research can therefore be formally stated as follows:

To test the feasibility of using DM to assess the relationship between and impact of KM on OR.

While this research relies on a questionnaire for data collection, the focus of this work is not on the data itself; rather, it focuses on determining if DM tools are able to capture the intricate relationships that exist between many knowledge management processes and organizational resilience. For this reason, a number of DM techniques are tested, with each involving the building and testing of a DM model using resilience as the independent variable and KM processes as the dependent variables.

Because of the researcher's professional affiliation and the research interests, the focus of the research is on SMBs, which are defined in Chapter 4. However, the applicability of the research findings is not limited to SMBs.

1.3 Research Questions

Table 1.3.1, below, presents the research questions and objectives that support the aims of this research:

Research question #1: What prior research exists regarding the application of DM with respect to KM and OR and the impact of KM on OR and what are the known relationships between KM and OR?	Objective: To determine the feasibility of using DM when evaluating KM, OR and/or the impact of KM on OR. Also, to determine the applications of DM techniques that have been developed in support of KM and OR as well as to identify the areas of convergence between DM, KM and OR.
Research question #2: Can OR be measured pragmatically? Can the impact of KM on OR be pragmatically measured?	Objective: To determine if OR (as defined in this research) can be measured. Also, to determine if the impact of KM on OR can be measured and how previous attempts to make such measurements can inform this research. In addition, the findings are to be used in formulating the OR section of the questionnaire used in the research.
Research question #3:	Objective:

Which KM processes are the most influential in achieving OR?	To explore the uses of DM in order to test its suitability for assessing the primary grouped data provided by the questionnaire answers, with the purpose of identifying their relationship with OR.
Research question #4: Can a methodological approach be developed to examine the relationships between KM and OR, utilizing DM?	Objective: To develop and apply a DM-based methodological approach for the analysis of data gathered from the use of the questionnaire and the generation of valid findings for this research.
Research question #5: Which are some of the main challenges encountered when employing DM for the purpose of determining the impact of KM on OR?	Objective: To identify the main issues (data, algorithm, error, algorithm parameters) associated with the use of DM for the purpose of measuring the impact of KM on OR.

Table 1.3.1: Aim, objectives and research questions

1.4 Structure of this Thesis

The structure of this thesis is presented graphically in Fig. 1.4.1.

Following an introduction to this thesis' research topic and objectives in Chapter 1, Chapter 2 presents the basic concepts and definitions used throughout this work.

The literature review presented in Chapter 3 has been divided into sections corresponding to the reviewed area. It includes knowledge management, organizational resilience, and a review of the utilization of data mining with respect to knowledge management and organizational resilience, the review seeking to determine the impact of knowledge management on organizational resilience through the data mining lens.

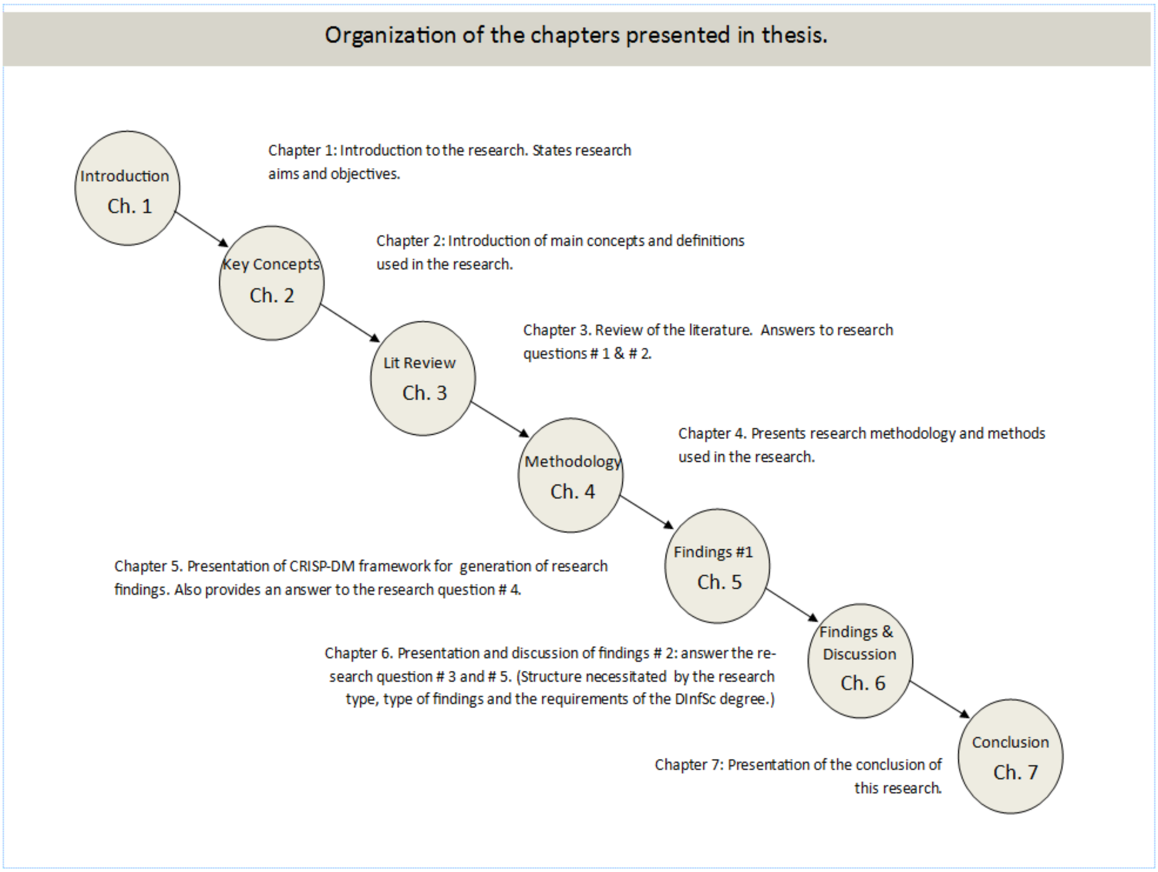
The methodology used in this research is presented in Chapter 4.

The first part of the findings – the actual application of the data mining models with respect to the measurement of the impact of knowledge management on organizational resilience – is presented in Chapter 5. Chapter 5 also presents a methodological approach for conducting data mining projects that seek to investigate the impact of knowledge management on organizational resilience.

The second part of the findings, along with a discussion thereof, is presented in Chapter 6. (Given the research approach used, which focuses on methodology rather than numerical output, and as each data mining model is discussed individually, separating the discussion from the findings made the former seem fragmented and harder to follow; hence, it was decided to add the discussion to the findings presented in the same chapter.)

Chapter 7 is the concluding chapter of this work.

In addition, given the applied nature of the research and the large number of data mining models presented, there are number of supporting appendices.


Organization of the chapters presented in thesis.

Chapter 1: Introduction to the research. States research aims and objectives.

Chapter 2: Introduction of main concepts and definitions used in the research.

Chapter 3. Review of the literature. Answers to research questions # 1 & # 2.

Chapter 4. Presents research methodology and methods used in the research.

Chapter 5. Presentation of CRISP-DM framework for generation of research findings. Also provides an answer to the research question # 4.

Chapter 6. Presentation and discussion of findings # 2: answer the research question # 3 and # 5. (Structure necessitated by the research type, type of findings and the requirements of the DInfSc degree.)

Chapter 7: Presentation of the conclusion of this research.

Fig. 1.4.1: The structure of this thesis

1.5 Summary

This chapter briefly introduced this thesis in terms of its research problem, the context of and justification for its research and its general layout. The next chapter introduces key terms and definitions, providing the foundation for the chapters that follow.

CHAPTER TWO: DEFINITION OF KEY CONCEPTS

2.1 Introduction

This chapter introduces and defines concepts critical to the topics addressed within this thesis. The key concepts used in this thesis and defined in this chapter include the following: knowledge, a concept that is introduced in the context of knowledge management in order to facilitate discussion of the actual management of knowledge; knowledge management, a key aspect of this thesis, which focuses on the management of knowledge in the business setting; organizational resilience, a key concept in the business context studied in this thesis; and data mining, the analytical instrument used in this research, the basic aspects of which are introduced. The topics covered in this chapter are illustrated in Fig. 2.1.1, below:

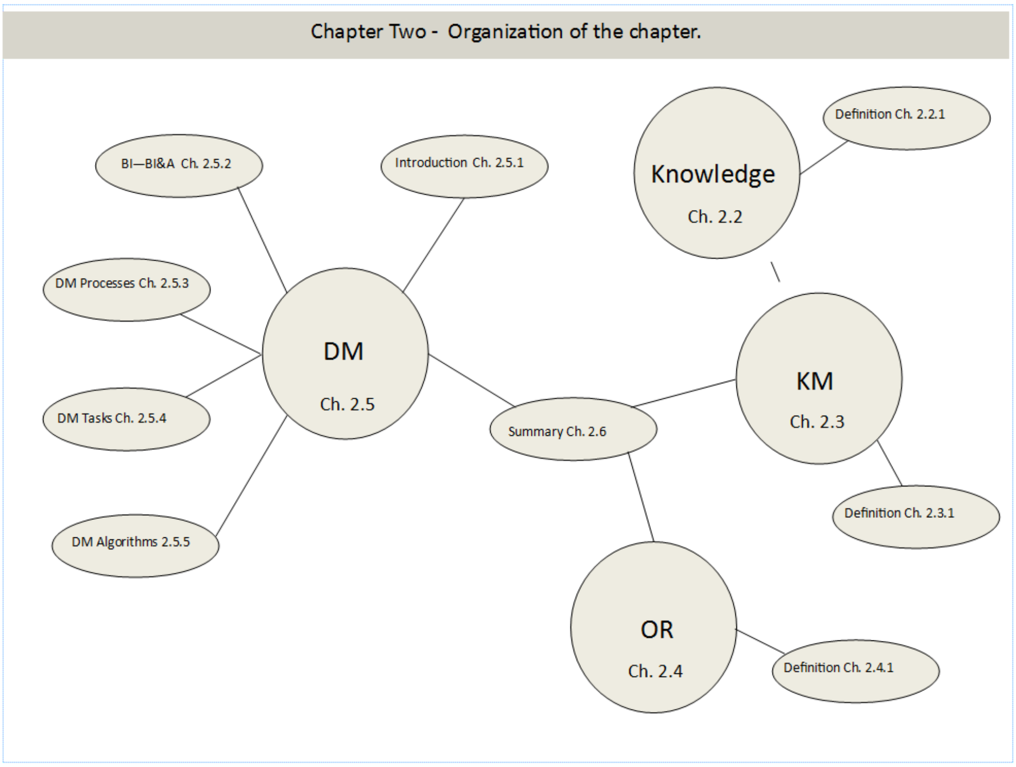


Fig. 2.1.1: Representation of topics covered in Chapter 2

2.2 Knowledge

2.2.1 Definition of Knowledge

Prior to defining knowledge itself, an introduction to the various views of knowledge is required. Over the years, numerous views of knowledge itself have been proposed. Some of the main views have been discussed by Alavi and Leidner (2001, pg. 111) and include the following:

- The hierarchical view of knowledge: Data consists of facts and raw numbers. Information is processed/interpreted data. Knowledge is personalized information;
- State of mind: Knowledge is the state of knowing and understanding;
- Object: Knowledge is viewed as an object to be stored and manipulated;
- Process: Knowledge is a process that consists of applying expertise;
- Access: Knowledge is a condition in which access to information is possible;
- Capability: Knowledge is seen as having the potential to be used to influence an action; and
- Holder of knowledge: Knowledge is viewed as existing in the individual or the collective.

While numerous definitions of knowledge exist, the definition of knowledge used for the purpose of this research comes from the work of Bergeron (2003, pg. 11) and is presented, in hierarchical form, in Fig. 2.2.1.1, along with the elements related to knowledge. (The definition of the term knowledge is provided on the following page, which discusses the knowledge elements of the hierarchy.) This definition has been chosen to clearly illustrate the impact of data, information and metadata on knowledge itself, and it lends itself very well to this work, which also begins with the data layer and, through the use of data mining, moves towards knowledge. Moreover, the clear separation between the computer and the machine provides an additional dimension for considering the role of human and computer in this research, which is also mentioned in Chapter 3.4's discussion of the DM field. When discussing the management of knowledge, Bergeron's model, in addition to knowledge management, also supports the approach of managing layers to achieve knowledge, making the management aspect more robust.

The hierarchy-based view of knowledge (Bergeron, 2003, pg. 11) provides an explanation of the terms used in this research – 'data', 'information' and 'knowledge' – which are commonly misused in practice.

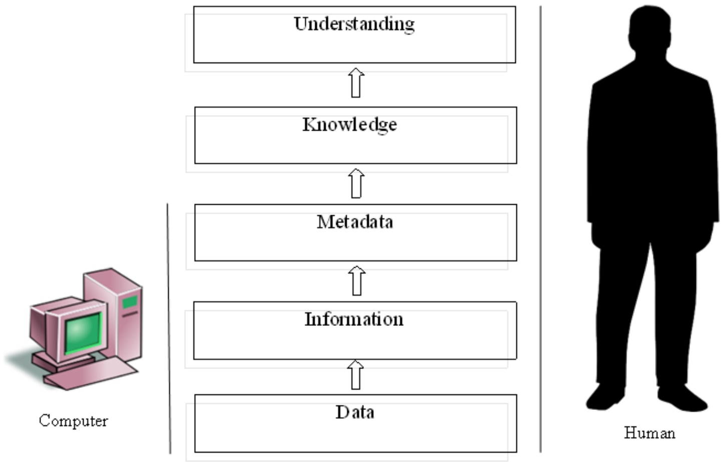


Fig. 2.2.1.1: Hierarchical presentation of knowledge [Derived from Bergeron (2003, pg. 11).]

Each component of the knowledge hierarchy shown in Fig. 2.2.1.1 is defined by Bergeron (2003, pg. 10) as follows:

'**Data** are numbers. They are numerical quantities or other attributes derived from observation, experiment, or calculation.'

'**Information** is data in context. Information is a collection of data and associated explanations, interpretations, and other textual material concerning a particular object, event, or process.'

'**Metadata** is data about information. Metadata includes descriptive summaries and high-level categorization of data and information. That is, metadata is information about the context in which information is used.'

'**Knowledge** is information that is organized, synthesized, or summarized to enhance the comprehension, awareness, or understanding. That is, knowledge is a combination of metadata and an awareness of the context in which metadata can be applied successfully.'

‘Instrumental **understanding** is the clear and complete idea of the nature, significance, or explanation of something. It is a personal, internal power to render experience intelligible by relating specific knowledge or broad concepts.’ The highest level of the hierarchy, understanding, is outside the focus of this work and has been defined here solely to provide a complete account of the hierarchy.

The data-information-knowledge-wisdom (DIKW) hierarchical model, as it is referred to by Fricke (2007), Weinberger (2010) and Smith (2011), which is also the classification for the model presented by Bergeron (2003), is not without its critics. Smith (2011, pg. 2) lists Fricke (2007) as the key critic of the model. Smith states that Fricke’s critique is ‘[b]ased on outmoded metaphysics of materialism, positivism’ but goes on to use the model, stating it is applicable for the needs of his work. Inspecting the work of Fricke (2007, pgs. 10-11), some of the key issues identified by the author include statements such as the following: ‘All data is information. However, there is information that is not data.’ This statement is not of great significance for this thesis, as Bergeron’s model, which is used as the knowledge model in this research, explicitly depicts the direction of flow as being only from data to information, not vice versa. Fricke’s other concern, regarding the DIKW model seeking knowledge in the form of ‘know-how’ (how to ride a bicycle, for example) instead of ‘know-that’ (knowing that Aberdeen is in Scotland, for example), is beyond the scope of this research. As a matter of fact, the author of this work would be fully satisfied if its research results simply addressed ‘know-how’ knowledge, but the work makes no distinction between the kind of knowledge the hierarchical knowledge model is to hold. With this in mind, this thesis uses Bergeron’s hierarchical knowledge model.

2.3 Knowledge Management

As an extensive discussion of knowledge management (KM) (building on the concept of knowledge just discussed) is provided in Chapter 3.2, this section focuses on defining KM.

2.3.1 Definition of Knowledge Management

As a multidisciplinary field, knowledge management can be defined in a number of ways. In addition, the large number of possible definitions of the term knowledge presented in the previous section leads to a variety of different perceptions of KM. To illustrate the variations between various definitions of KM, some of the key definitions are presented below.

Interestingly, one of the key early pioneers in the KM field, Sveiby, did not personally like the name of knowledge management for the field: 'Personally, I dislike the notion of KM. Knowledge is a human faculty, not something that can be managed, except by the individual him/herself. A better guidance for our thinking is therefore phrases such as "to be Knowledge Focused" or to "see" the world from a "Knowledge Perspective". To [Sveiby] KM is 'The Art of Creating Value from Intangible Assets'" (1996, pg. 1).

This view is further supported by von Krogh, quoted by Alavi and Leidner (2001, pg. 113) as stating that 'KM refers to identifying and leveraging the collective knowledge in an organization to help the organization compete.'

A very different definition of knowledge, and therefore a different view of KM, is provided by Wilson (2002, 2), a key critic of KM, who states that '[k]nowledge is defined as what we know: knowledge involves the mental processes of comprehension, understanding and learning that go on in the mind and only in the mind, however much they involve interaction with the world outside the mind, and interaction with others'. In his work, Wilson (2002) provides a unique view of KM, primarily from the ecological perspective. According to him and based on his view of knowledge (as introduced in the section above), knowledge is what a person knows and resides only in that person's mind. On the other hand, Wilson refers to information as something that does not exist in the brain but rather outside of it (in books and databases, for example). Thus, according to Wilson, in order for information to become knowledge, it must be absorbed by the knowledge structures that exist within a person.

Based on this, Wilson states that KM is a management fad that rests on two pillars: the management of information and the effective management of work practices (2002, pg. 20). Wilson's argument regarding the definitions of knowledge, KM and information and the assumptions he uses would benefit from further review for two reasons: first, the key components of the definition he uses are not well defined and, second, if knowledge resides only in an individual, then organizations would have no knowledge if individuals leave; i.e., there would be no organizational knowledge. Clearly, some knowledge within an organization could constitute work practice (provided that work practice is defined) but not all knowledge. Some of the knowledge generated, for example, from the application of business intelligence, may or may not be classified as work practice. Wilson also supports Sveiby's view that knowledge cannot be managed, a view that holds KM as something that belongs to an individual, rather than an organization.

In addition to the difficulties of defining KM at the personal level, among the writers who support the organizational view of KM, the categorizations and definitions of KM are also not uniform.

The difficulty in determining a definition of KM has been noted relatively recently by Frappaolo (2006, pg. 8). Frappaolo writes that, KM is not a matter of technology, strategic directive, business strategy or culture alone; they should all be considered in KM. According to Frappaolo, 'KM is the leveraging of collective wisdom to increase responsiveness and innovation.' Per this definition, KM should result in a positive outcome for an organization.

Carlucci and Schiuma (2006, pg. 36), based on their literature review, recognize two main characteristics of KM that are supported by the definitions of KM used by various researchers. The first characteristic deals with the management aspect of KM and represents the so-called (dynamic) process-view of KM (these processes include, for example, knowledge creation, sharing and dissemination). The other characteristic takes a resource-based view of knowledge and is more concerned with the organizational and static aspects of KM.

More recently, KM literature has devoted increased attention to the utilization of KM to benefit an organization (iJet International Inc., 2008, pg. 5; McCann et al., 2009, pg. 45; Wu et al., 2010, pg. 398); hence, the influence of such works on the definition of KM.

Wu et al. (2010, pg. 398) cite Benbya et al. (2004) and provide the following process-based definition of KM (which is also the approach used in this thesis, as will be shown in Chapter 3): 'KM is a systematic way to manage knowledge in the organizationally specified process of acquiring, organizing and communicating knowledge.' As stated by Wu et al. (2010, pg. 398), citing Kamara et al. (2002), 'KM is organizational optimization of knowledge to achieve enhanced performance through the use of various tools, processes, methods and techniques.'

To make the definition of KM even more interesting, Spender (2005, pg. 149) states that definitions of KM are not very important 'provided we do not stop theorizing before reaching a position that encompasses all three types of knowledge' , which, with some similarity to Bergeron's (2003) hierarchical presentation of knowledge, he identifies as knowledge-as-data, knowledge-as-meaning and knowledge-as-practice.

However, for the purpose of this work, both of the definitions of KM provided by Wu et al. in the prior paragraph will be used in order to emphasize two main aspects considered in this research: a systematic way of managing knowledge and the utilization of KM for the improvement of an organization.

2.4 Organizational Resilience

As the extensive discussion of organizational resilience (OR) is provided in Chapter 3.3, this section focuses on defining OR.

2.4.1 Definition of Organizational Resilience

Traditionally, OR was understood to refer to crisis management, being the ability to survive a tragic, single, event, but the field grew to include divergent meanings drawn from the fields of business, medicine, psychology, ecology and economics, to name but a few. As stated by Ponis and Koronis (2012, pg. 923), who undertook an extensive, peer-reviewed literature review of OR-related texts, their '[l]iterature study proves the existence of two discrete approaches on organizational resilience. Some scholars see organizational resilience as a simply an ability to rebound from unexpected, stressful, adverse situations and pick up where they left off, while others visualize organizational resilience beyond restoration to include the development of new capabilities and an expanded to keep pace with and even create new opportunities', with the later approach being the research topic of this dissertation.

For the purpose of this document, if not stated otherwise, OR refers to the business domain and one of the many interpretations of the term used in business: OR is the ability of an organization to remain in business (and perhaps even flourish) under adverse business conditions.

Over the years, the concept of OR has been defined in many ways (Mallak, 1998; Robb, 2003; Hamel & Valinkagas, 2003; iJet International Inc., 2008; Braes & Brooks, 2010), influenced by social, political and economic forces. The common quality in all of the OR definitions identified by OR researchers and practitioners, considering the 'non-crisis-based view of OR', is the need to be able to sense and adjust to changes in the business environment (Robb, 2003; Hamel & Valinkagas, 2003; McCann et al., 2009). In addition, the need for enabling OR factors, such as organizational structure and culture, is highlighted by another group of scholars (Horne & Orr (1998); McCann et al. (2009); Cockram & van Del Heuvel (2012)). Finally, some scholars view OR as being based on engineering studies (Horne (1998); Mallak (1998); Robb (2000)), while others see it as based on natural and/or biological studies (Sundstrom & Hollangel (2006), Friedman (2005); Coutu (2002)). Taking a chronological perspective, from the earliest attempts to define OR, it can be seen that, according to Horne (1997, pg. 27), 'Organizational resilience is the ability of a system to withstand the stresses of environmental loading based on the combination/composition of the system pieces, their structural inter-linkages, and the way environmental change is transmitted and spread throughout the entire system. To varying degrees, resilience is a fundamental quality found in individual, groups, organizations, and systems as a whole. It allows a positive response to significant change that disrupts the expected pattern of events without resulting in regressive/nonproductive behavior.'

Mallak (1998, pg. 8) states that 'the resilient organization designs and implements effective actions to advance the organization, thereby increasing the probability of its own survival', and similarly to Horne, emphasizes the importance of individual resilience, with this emphasis being reflected in the resilience principles he proposes.

Robb's definition stems from a more systematic and balanced view of OR. Robb (2000, pg. 27) states that: 'A Resilient Organization is able to sustain competitive advantage over time through its capability to do two things simultaneously:

- Deliver excellent performance against current goals.
- Effectively innovate and adapt to rapid, turbulent changes in the markets and technologies.'

According to Hamel and Valinkagas (2003, pg. 2), resilience is 'the ability to dynamically reinvent business models and strategies as circumstances change.' Hamel and Valinkangas introduce the concept of 'strategic resilience', which refers to the use of a strategy that is constantly evolving and aligning itself to upcoming opportunities and current trends. They define strategic resilience, stating that it 'is not about responding to one time-time crisis. It's not about rebounding from a setback. It's about continuously anticipating and adjusting to deep, secular trends that can permanently impair the earning power of a core business. It's about having the capacity to change before the case for change becomes desperately obvious.' With strategy being a key component of organizational management, this description is also highly applicable to OR.

A paper presented by iJet Intelligent Risk Systems (iJet International Inc., 2008), a leading provider of global intelligence and business resiliency services, offers a definition of the term 'business resilience' that closely matches the definition of OR, being defined as 'the ability to rapidly adapt and respond to risks as well as opportunities in order to maintain continuity of business operations, remain a trusted partner and enable growth' (iJet International Inc., 2008, pg. 5). The paper reports that resilient organizations constantly monitor the world for changing threats and opportunities (e.g., risks, organizational changes and market changes) so that the negative impacts of destructive events can be avoided by acting appropriately before people and assets are affected (iJet International Inc., 2008, pg.5).

A very insightful definition and interpretation of OR is provided by Braes and Brooks (2010, pg. 14), who state that '[i]t is argued that Organizational Resilience is not an overarching philosophy, strategy, process or management system, but rather a foundation comprising the outcomes from many applied domains. Nevertheless, Organizational Resilience can be defined as a sum of essential concepts. These essential concepts include enterprise risk management, governance, quality assurance, information security, physical security, business continuity, culture and values supported by adaptive leadership.'

The review of existing OR-related work presented in Chapter 3.3 tends to validate the view/definition presented by Braes and Brooks, which states that OR is truly a multi-domain subject.

For the purpose of this research, the definition presented by Hamel and Valinkangas (2003) is used, as it concurs with the author's personal views regarding what OR is and what it takes for an organization to be resilient.

2.5 Data Mining

Due to the technical, as opposed to business-oriented, nature of data mining (hereafter referred to as DM), this concept is introduced on its own in this chapter as part of the background information of this thesis and because of the significant role of the DM models play within this research.

2.5.1 Introduction to Data Mining

Data mining, as stated by Aghdaie et al. (2014, pg. 768), 'is an interdisciplinary field that combines artificial intelligence, database management, data visualization, machine learning, mathematics algorithms, and statistics.' While there are numerous definitions of DM, the definition that is both most appropriate for the purpose of this thesis (and therefore used in it) and not overly verbose is that offered by Gullo (2015, pg. 18), who defines DM as 'the computational process of analyzing large amounts of data in order to extract patterns and useful information.' This definition captures the essence of the definition of knowledge presented in the previous section of this chapter and agrees very well with Bergeron's (2003, pg. 11) hierarchical definition of knowledge presented in Section 2.2.

In industries, DM is often used interchangeably with the concept of predictive analytics (PA), as the approaches and algorithms used by both disciplines are generally the same. Abbott, one of the authorities on DM/PA, states that (2014, pg. 13) 'I have treated the two fields as generally synonymous since predictive analytics became a popular term.' Abbott adds that there was a need for the new term as data mining received a great deal of negative publicity toward the middle of the first decade of the 21st century due to the Department of Defense and National Security Agency's widespread use of DM to analyze the communications of ordinary citizens.

To support Abbott's (2014) view concerning the similarities between PA and DM, the work of Chantal and Chantal should also be considered. Chantal and Chantal (2015, pg. 4) define PA as 'the process of extracting information from large data sets in order to make predictions and estimates about future outcomes.'

Traditionally, DM was performed on data sets generated as a result of data being collected for other reasons, such as capturing supermarket transactions that track how customers are billed for items purchased. As noted by Hand (2007, pg. 621), data sets are collected primarily for the purpose of data mining.

2.5.2 Business Intelligence (BI) – Business Intelligence & Analytics (BI&A)

Prior to the discussion of DM/BI/BI&A, a baseline definition of analytics is required. Analytics, as defined by Abbott (2014, pg. 2), 'is the process of using computational methods to discover and report influential patterns in data.' (As can be already seen, this is very similar to the definition of DM provided by Gullo presented in the previous section. This is one of the examples of the ambiguity that is possible when two distinct fields are defined similarly; therefore, further clarification of terms that are often used interchangeably is presented below.)

As stated in the section below and shown in the literature review, the terms DM, analytics and business intelligence (BI) are often used interchangeably.

While the treatment of the terms as similar is appropriate for this work, the terms must be defined and the differences and similarities between them outlined in order to clarify their roles and relationships within this research.

The definition of BI provided by Watson (2009, pg. 6) – 'a broad category of applications, technologies, and processes for gathering, storing, accessing, analyzing data to help business users make better decisions' – includes DM because DM, both in its most general form and within the business context, enables and facilitates decision-making. Watson's definition also includes all of the preparatory steps that deal with data-loading and data-cleaning. The concepts and processes involved in DM are described in the next section. Because of its comprehensive nature, the definition of BI offered by Watson is the definition chosen for this work; this is also the definition used in the work of Isik et al. (2013, pg. 13).

Similarly, the practitioner-based BI definition provided by Larson (2009, pg.11) also views BI as a governing concept for data mining, analysis and decision-making: 'Business Intelligence is the delivery of accurate, useful information to the appropriate decision makers within the necessary timeframe to support effective decision making'.

Regarding BI&A, Kowalczyk et al. (2013, pg. 3) cite Davenport (2010) and Watson (2010) and refer to BI&A as 'includ[ing] collection, analysis and dissemination of information with the purpose of supporting decision making.'

Seeing BI as a support platform for business decisions (Turban et al., 2007; Watson 2010; Larson, 2009) allows analytics, DM and PA to be perceived as tools used to specifically supporting such a platform. Thus, for the purpose of this research, Fig. 2.3.2.1 represents the assumed interrelationships between BI, BI&A, DM and PA, in which the following observations should be borne in mind:

- Business intelligence is synonymous with business intelligence and analytics;
- Data mining is synonymous with predictive analytics; and
- Data mining is a component of the broader concept of business intelligence.

In alignment with the definitions and concepts discussed in the literature review, for the purpose of this work, the terms BI, BI&A, DM and PA may be used interchangeably, unless otherwise noted. However, to facilitate comprehension of the concepts and relationships between these terms, Fig. 2.3.2.1 is presented below, wherein BI and BI&A form the superset of the topics/functionalities addressed by DM and PA.

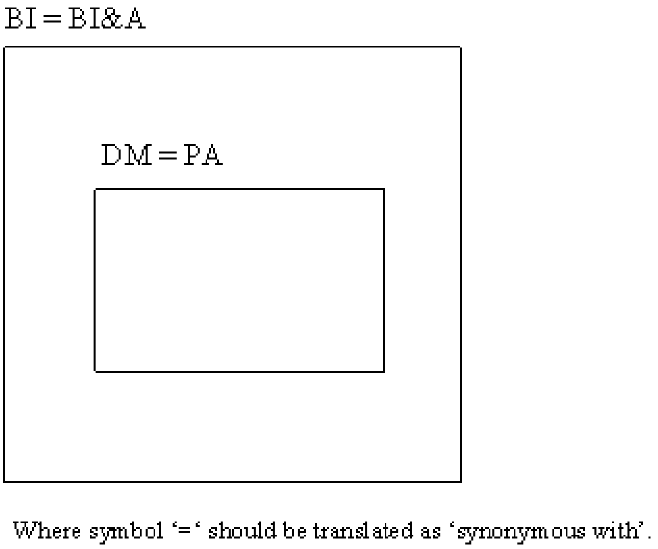


Fig. 2.5.2.1: DM and PA as a component of BI and BI&A

2.5.3 The Data Mining Process

Data mining encompasses analytical tools and algorithms as well as processes. One of the methodologies that is widely used in the field and is independent of the underlying data mining algorithm used is the Cross Industry Standard Process for Data Mining (CRISP-DM), which was originally released in the 1990s. The industry standard CRISP-DM model allows for a comprehensive and methodological approach to DM, ensuring that the key aspects of DM are carried out and that they are performed in a specific order. For this reason, this thesis uses the CRISP-DM model. Individual stages of the model are discussed in detail in Chapters 4, 5 and 6.

An example of research work utilizing data mining as the knowledge discovery tool in analysis of school performance and CRISP-DM model was the work of Alsutanny (2011).

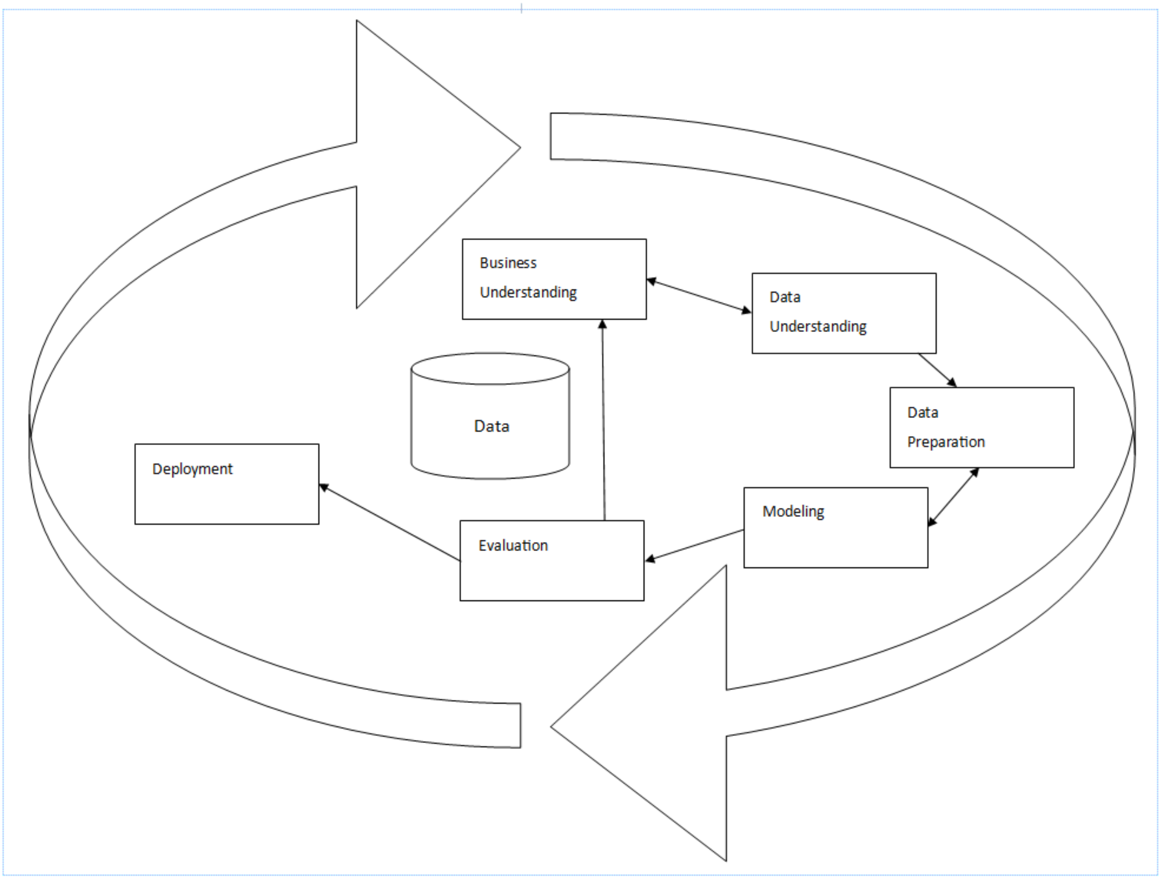


Fig. 2.5.3.1: CRISP-DM. [Derived from IBM (SPSS, 2000).]

2.5.4 Tasks Accomplished by Data Mining

According to Witten et al. (2011, pg. 8), data mining constitutes practical, non-theoretical learning that uses techniques for finding and describing structural patterns in data for the purpose of explaining data and making predictions. The most common data mining tasks (MacLennan et al., 2009, pg. 6) include the following:

- Classification – the act of assigning a category to each case investigated. ('Each case contains a set of attributes, one of which is the class attribute. The task requires finding a model that describes the class attribute as a function of input attributes');

- Clustering – also known as segmentation. Clustering is used to identify natural groupings of cases based on the set of attributes. (Cases within the same group tend to have similar attribute values);
- Association – also known as market basket analysis. Association seeks to identify items that frequently appear together (for example, in a sales transaction) and then, based on this information, determine the rules about associations between items;
- Regression – similar to classification; however, rather than searching for patterns that describe a class, the goal is to find patterns that determine numerical value;
- Forecasting – takes as an input a sequence of numbers that indicates a series of values through time and then computes the future values of that series;
- Sequence analysis – finds patterns in a series of events (such as browsing through a web site); and
- Deviation analysis – finds cases that behave very differently from the norm.

In addition to the tasks identified by MacLennan, Jackson (2002, pg. 276) also notes another very important task:

- Dependency analysis – used to predict the value of an item given information about other items.

Each DM task is supported by one or more DM algorithms, where the DM algorithm is an automated extraction of data patterns that are applied to data and includes techniques such as decision trees, Naïve Bayes, time series and neural networks. (For the purpose of this discussion, as it occurs in the field, the terms 'DM algorithm' and 'DM technique' are used interchangeably.) The output of an algorithm is a set of rules, called a mining model, that describes the effects of changing one or more variables on another variable or set of variables (Janus & Misner, 20111, pg. 347).

2.5.5 Data Mining Algorithms

There are nine DM algorithms available within the Microsoft's SQL Server 2012, which are listed below. While all of the algorithms are listed here to ensure that key concepts are presented in full, the algorithms applicable to this research are discussed in Chapter 6. The available algorithms include the following:

- Naïve Bayes
- Decision trees
- Microsoft Linear Regression
- Microsoft Logistic Regression
- Microsoft Neural Network
- Microsoft Clustering
- Microsoft Sequence Clustering
- Microsoft Time Series
- Microsoft Association Rules

2.5.6 Domain Driven Data Mining

One of the latest developments in the DM field, referred to by Zhang et al. (2010, pg. 753) as the 'next-generation data mining framework', is domain-driven data mining (DDDM), which originated from the realization that data mining needs to have context for both defining the problem and interpreting results. Only examining the data, without taking into consideration domain factors, appeared to not deliver the payoff expected from DM initiatives.

As stated by Zhang et al. (2010, pg. 753), the aim of DDDM is to embed the domain-related factors and synthesized ubiquitous intelligences affecting the domain of the problem with the knowledge discovery that results from the application of DM algorithms. These actions, as pointed out by Zhang et al., are based on domain-expert knowledge, constraints, organizational factors, domain adaptation and operational knowledge.

Kumari (2011, pg. 2) states that 'Domain Driven Data Mining is proposed as a methodology and a collection of techniques targeting domain driven actionable knowledge delivery to drive Knowledge Discovery from Data toward enhanced problem-solving infrastructure and capabilities in real business state of affairs.'

While the DDDM framework does provide more focus on DM, and therefore greater anticipation of positive impactful results, the framework is not easily applicable to the cross-industry data mining; rather, it can only be applied within an individual organization. The reason for this limited scope of application is the application of domain experts (many times from within an organization), organization specific constraints, organizational factors and operational knowledge. Given these limitations and the fact that this thesis seeks to develop a tool that can be applied across industries, the DDDM framework is not used in this research; it is mentioned here only to ensure that the literature review is thorough.

2.6 Summary

This chapter has presented concepts that are key to the topics addressed within this thesis and the relationships between these concepts; its content is critical in understanding the nature of this research and its significance. Chapter 3, which follows, discusses the findings of the literature review with respect to KM, OR and the impact of DM on KM and OR.

CHAPTER THREE: REVIEW OF THE LITERATURE

3.1 Introduction

The chapter aims to identify and examine relevant works that have already been conducted in the areas involved in this research, determine their importance in relation to this thesis' research, identify the nature of the research problem, identify the major factors involved in the problem, and highlight the gaps in theory and practice that were identified as a result of the review. The purpose of the review is to develop an understanding not only of the impact of the literature within the involved disciplines but also of the relationships that exist between these areas. This chapter also addresses the first two research questions by answering the following question: What prior research exists in the areas relevant to this thesis? Based on the literature review and the gaps identified in it, Section 3.5 introduces a new theoretical model that builds on the findings of the literature review and the gaps identified. The summary section (Section 3.6) provides an overall summary of Chapter 3. The high-level layout of Chapter 3 is presented in Fig. 3.1.1, below:

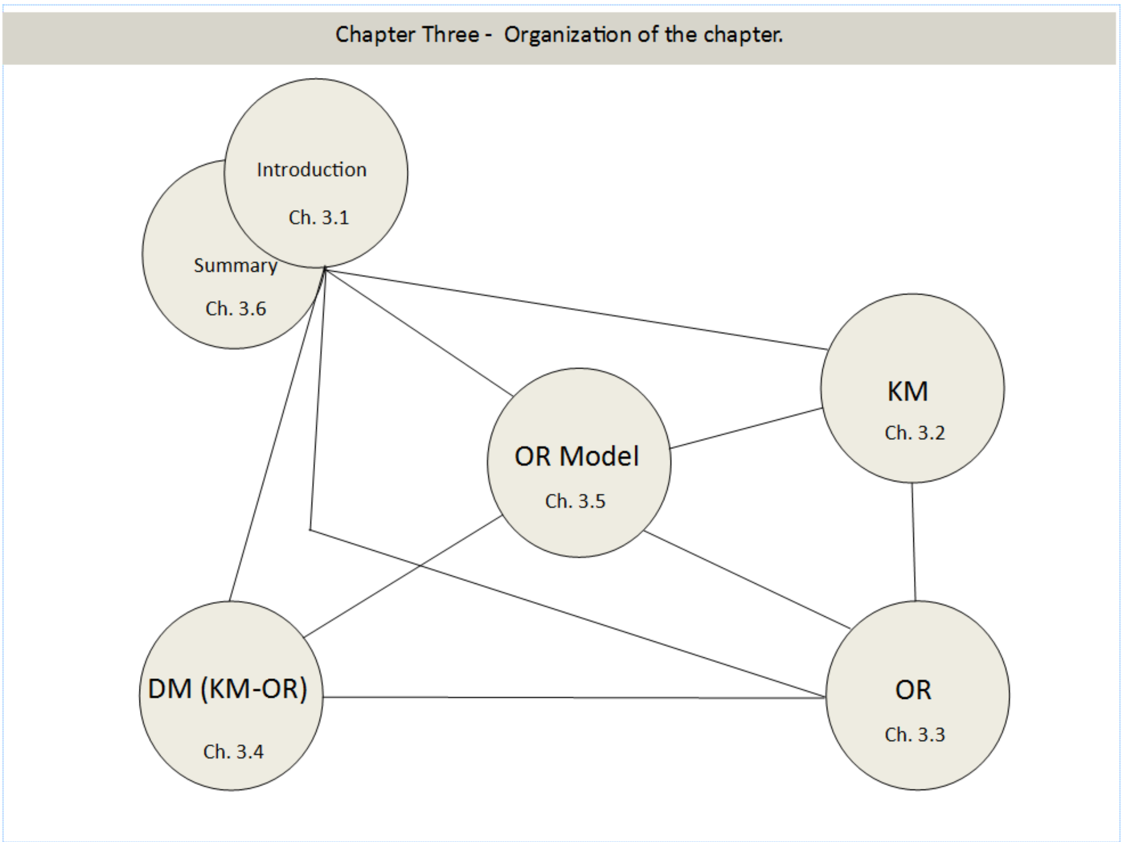


Fig. 3.1.1: The high level organization of Chapter 3

3.2 Knowledge Management

3.2.1 Introduction

This chapter builds on the concepts and definitions of knowledge and knowledge management introduced in Chapter 2. The foundational concepts in the area of what is known today as knowledge management come, to a large extent, from the work of Polanyi (1966; 1974).

Polanyi was the first writer who considered the concept of tacit knowledge which, very generally speaking, can be described as the hard-to-articulate knowledge that resides within us. Polanyi wrote that 'we can know more than we can tell' (1966, pg. 4). The field of KM then had its beginning as a formal discipline with the work of Nonaka and Takeuchi (1995) and has been constantly evolving since. The KM field draws from a number of disciplines, including business administration, information systems and management, library and information sciences (Alavi & Leidner, 1999).

Due to the influence of various disciplines on the KM field, there are a number of possible ways in which the schools of thought within the literature can be grouped. Some writers, especially those from the earlier period, tend to divide the KM field into the following categories: techno-based, with a focus on technology (Horne (1997); Malhorta (1998); Mallak (1998); and Frappaolo (1998)); organization-based, with a focus on how organizations can be designed to promote KM (Nonaka (1991) and Hussain (2004)); and ecologically based, with a focus on people, their interactions and environmental systems (Nonanka (1991); Horne (1997); Mallak (1998); Gupta & McDaniel (2002); Murray (2002); McElroy (2003); Hamel & Valinkangas (2003); McKenzie & van Winkelen (2004); and McCann et al. (2009)). In addition to the groupings based on the KM focus, there are a number of KM strategies, including, among others rewards, storytelling (Gabriel, 2000), communities of practice (Wenger, 1998), knowledge repositories (Liebowitz, 1999), and best practices (Szulanski, 1996). Finally, there are a number of proposed theoretical KM frameworks (which are further discussed in Section 3.2.3): Demerest's KM model (McAdam & McCreedy, 1999), Frid's KM model (Frid, 2003), Stankosky and Baldanza's KM framework (Stankosky & Baldanza, 2001), Kogut & Zander's KM management model (Kogut & Zander, 1992), McElroy's knowledge lifecycle model (Haslinda & Sarinah, 2009), and McKenzie and van Winkelen's competence model (McKenzie & van Winkelen, 2004, pg. 3). Of these models, the ones that receive the most attention in the KM field are discussed later in this chapter. Given the practical nature of this work, the discussion would not be complete without a review of the literature with respect to the role and value of KM in organizations, as the KM literature review focuses primarily on the application of KM in a business environment. Section 3.2 closes the literature review by examining KM's role in business and business value and the impact of KM on OR. A graphical presentation of the contents of Section 3.2 is shown below, in Fig. 3.2.1.1:

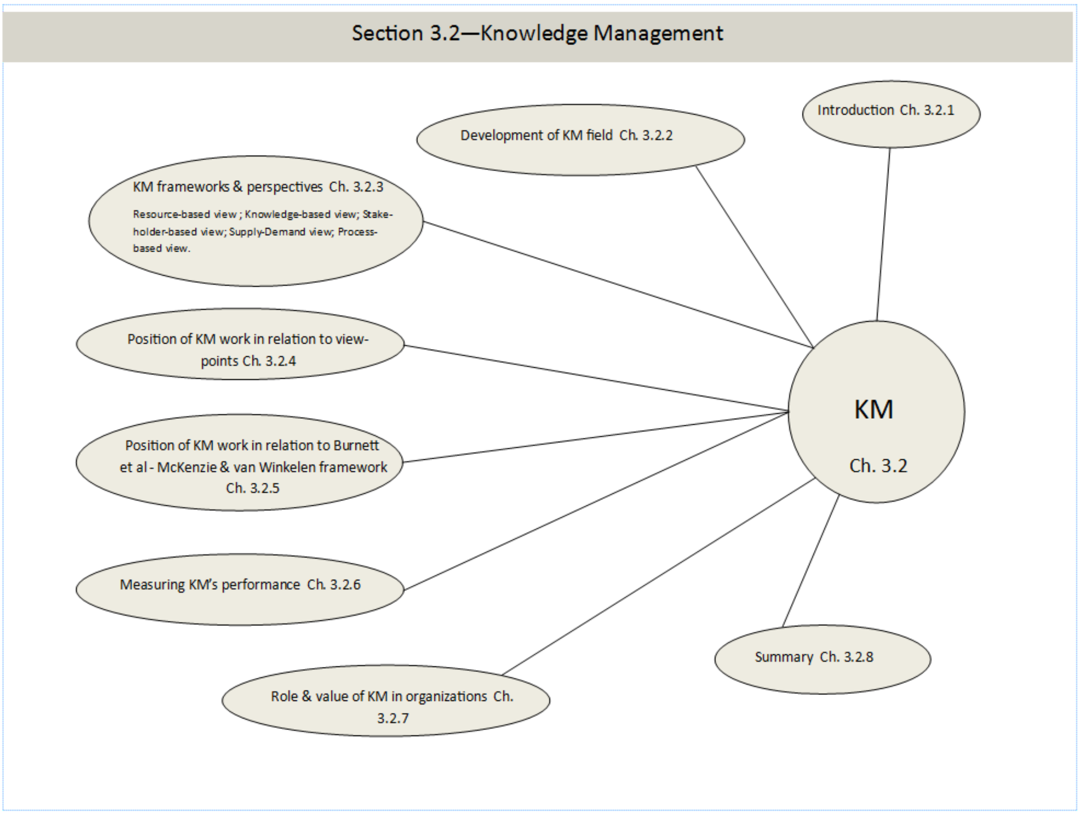

Section 3.2—Knowledge Management
Introduction Ch. 3.2.1
Development of KM field Ch. 3.2.2
KM frameworks & perspectives Ch. 3.2.3
Resource-based view ; Knowledge-based view; Stake-holder-based view; Supply-Demand view; Process-based view.
Position of KM work in relation to view-points Ch. 3.2.4
Position of KM work in relation to Burnett et al - McKenzie & van Winkelen framework Ch. 3.2.5
Measuring KM's performance Ch. 3.2.6
Role & value of KM in organizations Ch. 3.2.7
Summary Ch. 3.2.8
KM
Ch. 3.2

Fig. 3.2.1.1: Graphical representation of the contents of Section 3.2

3.2.2 The Development of the Knowledge Management Field

Prior to focusing on very specific aspects of the KM field for the purpose of this research, the literature review seeks to develop an appreciation of the development of the KM theories and the KM field. This section provides a summary of the historical views of the field as well as KM theories and applications that are important to this research.

One of the seminal works in the development of the field of KM comes from the work of Nonaka (1991) and Nonaka and Takeuchi (1995). Nonaka and Takeuchi's early work has been conducted in the context of Japanese companies that, at the time of his writing, were gaining significant competitive advantage in the marketplace; hence there was increased interest on the part of remaining players located outside of Japan. Based on the statement '[i]n an economy where the only certainty is uncertainty, the one sure source of lasting competitive advantage is knowledge' (Nonaka, 1991, pg. 96), Nonaka and Takeuchi (1995, pg. 73) presented a four-element knowledge creation model that attempted to respond to the constant marketplace changes and that was, in their view, largely responsible for the success of Japanese companies in the area of innovation. Such a view tends to support the view of knowledge as a major factor responsible for business performance.

The model presented by Nonaka and Takeuchi, rather than being deterministic, was based on the concept of spiral flow: new knowledge is constantly leveraged within an organization to reach new levels. The model included the concepts of tacit and explicit knowledge. Tacit knowledge, as stated by Nonaka, 'is highly personal. It is hard to formalize and therefore, difficult to communicate to others' (Nonaka, 1991, pg. 98). In addition, tacit knowledge has a very important cognitive dimension: 'It consists of mental models, beliefs and perspectives so ingrained that we take them for granted, and therefore, cannot easily articulate them' (Nonaka, 1991, pg. 98). Or, in perhaps oversimplified words, it is a combination of formal as well as informal knowledge further refined by a person's life experiences. Explicit knowledge, on the other hand, is knowledge that is easily shared and is contained in manuals, books, or other written documents.

Nonaka and Takeuchi's (1995) spiral model of knowledge creation falls under the mixture of the organizational approach and the ecological approach as his knowledge creation model relied on some aspects of organizational design, as well as it relied heavily on human interaction as a part of the four-phase knowledge creation model. The view of the organization, as presented by Nonaka (1991), Nonaka and Takeuchi (1995), did not reflect typical Western-style, mechanistic organization. The need for extensive human interaction within an organization is perhaps best illustrated by the tacit-to-explicit phase as, according to Nonaka (1991, pg. 99) 'to convert tacit knowledge into explicit knowledge means finding a way to express the inexpressible.'

The work of Nonaka and Takeuchi (1995), while widely accepted, did find a voice of criticism with Tsoukas (2002) being perhaps the strongest critic of their work. This criticism related foremost to the definition of the tacit aspect of knowledge and the knowledge conversion aspect: from tacit knowledge to explicit. Tsoukas (2002) suggests that Nonaka and Takeuchi's definition of tacit knowledge, differing from that of Polanyi (1996, pg. 4) 'ignores the essential ineffability of tacit knowledge, thus reducing it to what can be articulated' (2002, pg.15).

The work of Malhorta during the end of the 1990s took into consideration the synergy of technology (with technology being a very important topic at the time the author wrote) and behavioral issues as part of KM. In his view, KM is formed by the combination of technology and is mandatory in order to understand and react to changing business conditions. In addition, Malhorta (1998) expands the notion of KM, discussing it as a lens through which an organization views all of its processes. Malhorta's view of KM provides more breadth to the KM discipline by expanding the notion of KM as a lens through which an organization should view all of its processes. In addition, Malhorta's views of the changing business environment are shared by many organizational resilience researchers, indicating its importance as major factor (Thurow, 1996; Horne, 1997; Mallak, 1998; Hamel & Valinkangas, 2003; McCann et al., 2009). Malhorta's approach to understanding KM views it as a synergy between technology and its processing abilities and the human capacity for creativity and innovation.

Roughly contemporary with Malhorta's work was that of Frappaolo (1998), which builds, to a large extent, on the work of Nonaka and Takeuchi (1995). Frappaolo, in addition to discussing the notions of tacit and explicit knowledge, introduced the concept of implicit knowledge: knowledge that can be harvested from the owner to be codified.

The technological aspect of Frappaolo's work is also worth noting. Frappaolo recognized the need for the utilization of technology in his key KM applications while also emphasizing the need for human interaction, particularly in the cognition application area (which refers to the linking of knowledge to processes and the process of decision-making based on available knowledge).

Murray's (2002) research highlights a number of points about effective KM. This was one of the earliest works that considered the benefits derived from KM, a key topic in this thesis. Murray states (2002, pg. 70) that effective KM utilizes a top-down approach and is demand-driven: that is, KM starts by identifying at the desired business results; then, it considers actions that will produce the desired results and the knowledge needed to support these actions.

Murray also states (2002, pg. 70) that KM is not very effective in improving existing processes as those already contain KM and that KM is best used to obtain new capabilities. Murray also shares his view on technology as an enabler of KM but not the source of it; he emphasizes the role of people, stating 'performance only improves when people do things differently' (2002, pg. 77).

McElroy, in addition to his three-tier KM model (2003, pg. 10) which is composed of the Knowledge Management layer, the Knowledge Processing layer and the Business Processing layer and which explains the relationship between them, is also the inventor of 'The Knowledge Life Cycle (KLC)' model (2003, pg. 6). In his KLC model McElroy illustrates how Knowledge Production impacts Knowledge Integration that feeds Business Process Environment. It is worth mentioning that, as opposed to many models like software development's waterfall model, McElroy's model is not linear, rather it forms a loop by providing feedback out of the Business Processing Environment to Knowledge Production, therefore allowing for the validation of knowledge existing in the system as well as allowing for learning to take place. The work of McElroy (2003) allows placing KM in the context of business processes and those are the key factors in business organizations.

It appears that, by the year 2005, there was still significant confusion regarding what constitutes KM (Schlogl, 2005, pg. 8). The work of Schlogl attempted to clarify such confusion as well as to clearly distinguish between information management (IM) and KM. To support his argument, Schlogl provides an insightful map of IM (2005, pg. 3) that categorizes writers (with the categories consisting of management, information sciences, information systems and information management classics) based on the author's co-citation analysis of data from the Science Citation Index and the Social Citation Index. In conclusion, Schlogl identifies three major categories in the literature on information and KM: technology-oriented information management (primarily data management), content-oriented information management (the management of codified information) and KM; he also describes what types of publication fall under each category. It is worth noting that, of the three main categories, only KM appears to be focused on an organization's strategic aspects.

Vorakulpipat & Rezgui (2008, pg. 283) summarize the 'evolution path' of KM by stating that in order for a firm to be effective it needs to migrate from the knowledge sharing (what McElroy calls first-generation) to the knowledge creation culture (McElroy's second-generation) but also to move past that point and create 'sustained organizational and societal values', to which this research seeks to contribute.

The applied, business-focused nature of this research is well aligned with recent works that emphasize the role of KM in the creation of value for organizations. The following KM writers are representative of this trend in the literature: McKenzie & van Winkelen (2004), Carlucci & Schiuma (2006), Vorakulpipat & Rezgui (2008), Ibrahim & Reid (2009), West & Noel (2009), Vatafu (2011), Crook et al. (2011). The work of these writers is discussed in Section 3.2.7, with a focus on the role and value of KM in organizations. These writers are mentioned here for the sake of completeness.

3.2.3 KM-based Frameworks and Perspectives

In addition to discussing developments within the KM field, the literature review examined a number of KM models/frameworks and KM perspectives with the intention of identifying those suitable for the purpose of this research. This aspect of the review is investigated in this section.

In the selection of the models guiding this research, a number of models were considered. The following table presents the models that were considered but not chosen for the purpose of this research. (Note that it is not an exhaustive list of the models that were reviewed with regard to their suitability for this thesis' research; rather, this list presents models that focused more specifically on KM than simply knowledge). The primary guideline in the consideration of these models was their ability to properly capture the multidimensionality and complexity of KM – what Moayer and Gardner (2012, pg. 69) refer to as unstructured problems (those with intricate, non-linear relationships between dependent and independent variables). In addition to the model selected for this research (discussed in Section 3.2.5), the following table lists the models reviewed, providing a brief description of each model and why it was not selected for inclusion in this research:

Model:	Brief Description:	Reason(s) for Rejection:
Demerest's KM model (McAdam & McCreedy, 1999)	Emphasizes the construction of knowledge within an organization. Consist of four key processes: knowledge construction, knowledge embodiment, knowledge dissemination, and knowledge use.	Model indicates directed, and therefore restrictive, flows. Model, due to number of processes considered, is inferior to that presented by Burnett et al. (2004; 2013).
Frid's (2003) KM model	Categorizes KM maturity levels and implementation in five levels: chaotic, knowledge-aware, knowledge-focused, knowledge-managed and knowledge-centric.	Appears to be more of a classification than comprehensive KM model.
Stankosky and Baldanza's (2001) KM framework	Addresses the enabling factors, including learning, leadership, organizational culture and structure, and technological infrastructure.	While there is little doubt about the need for the enabling factors, this research required a more comprehensive model that addressed KM processes.
Kogut and Zander's (1992) KM management model	Consists of five KM processes: knowledge creation, knowledge transfer, processing and transformation of knowledge, knowledge capabilities and individual "unsocial	While the model is slightly inferior to that presented by Burnett et al., it appears to contain most of the KM processes commonly mentioned in the KM literature. The lack of

	sociality"	explicit knowledge application and exploitation was a major weakness.
McElroy's (2003) knowledge lifecycle model	Integrates demand-side and supply-side KM with the integrated feedback component, making the model highly adaptive.	The model's processes exclude knowledge application and exploitation, a key KM component that affects this research.

Table 3.2.3.1: Some of KM models considered in this research

In addition to the selection of a model suitable for this research, the key KM perspectives were reviewed and considered, and the perspective most appropriate for this work was chosen (this is further described in Section 3.2.3.1). In addition, some of the views presented in the section below are a part of the theoretical OR model introduced in Section 3.5.

Resource-based view of KM

The resource-based theory (RBT), introduced by Barney, Lippman and Rumelt (Crook et al., 2011, pg. 444), views human capital (knowledge, skills, and abilities) as a resource that can lead to sustainable competitive advantage (or, at least, an advantage that lasts for a long time). Moreover, the RBT views human capital as a hard-to-replicate and not readily available resource that is semi-permanently tied to a firm and distinguishes it from similar organizations.

Knowledge-based view of KM

The knowledge-based view (KBV) is a perspective that emerged from the RBT and argues that knowledge embedded within people is ultimately the only source of competitive advantages (sf. Grant, 1996). Chou (2011, pg. 1594) states that the 'knowledge-based view of a firm suggests that knowledge is one of the most important resources of the firm and hypothesizes the objective of the firm is to integrate and create valuable knowledge.'

Stakeholder-based view of KM

The work of Moayer and Gardner (2012, pg. 69), citing Freeman and McVea (2001) and Gardner's (2001) political perspective or stakeholder-based view (SBV), highlights the importance for organizations of working with constituents or shareholders in order to achieve business goals and create competitive advantages. The constituents, per Freeman (2010, pg.42), are management, the local community, customers, employees, suppliers and owners. The political perspective addresses the need for a political process that identifies, classifies and cultivates positive relationships with stakeholders.

Supply-side vs. demand-side view of KM

Another method of categorizing KM approaches found in relatively recent literature is that of the classification into demand-side or supply-side. A description presented by McElroy (2003, pg. 14) provides an excellent explanation of the meaning of the term 'supply-side': 'KM interventions aimed solely at the enhancement of knowledge sharing, or integration, can be thought of as "supply-side" in their orientation because of their focus on enhancing the supply of existing knowledge.'

The demand-side, according to McElroy, is different: 'Practitioners of demand-side KM are mainly interested in enhancing an organization's capacity to satisfy demands for new knowledge' (2003, pg. 14).

Interestingly, according to McElroy, the supply-side characterizes what he refers to as the 'first generation' of KM, whereas an emphasis on both demand-side as well as supply-side characterizes the 'second-generation' KM (2003, pg. 14). This recognizes that both knowledge sharing and knowledge creation are critical to KM, which is a view shared by the author of this thesis.

3.2.3.1 Process-based View of KM

Alavi and Leidner's work (2001) in the area of KM led them to note that KM is largely viewed from a process-based perspective that involves various activities. An investigation of the KM literature reveals some basic KM processes that appear in many KM writings. Liebowitz (1999) identifies a number of process models proposed by different authors in the field, all of which consist of varying numbers of 'steps'. DiBella and Nevis (1998) suggest the simplest, a three-phase model: acquire, disseminate, and utilize. A number of authors suggest four-stage models. Wiig (1997), for example, suggests that KM consists of a four-stage process: creation and sourcing, compilation and transformation, dissemination and application and value realisation. Typically, basic KM activities include the activities of knowledge creation, knowledge transfer, knowledge storage and retrieval and knowledge application. Alavi and Leidner (2001, pg. 114), writing about KM processes/activities, state that '[s]light discrepancies in the delineation of the processes appear in the literature, namely in terms of the number and labeling of processes rather than the underlying concepts.' The purpose and design of such KM processes, as stated by Fink and Ploder (2007, pg. 705), are intended to ensure that an organization's profitability and competitive advantage in the marketplace are improved, which are key topics for this research.

This thesis builds on the process-based view of a firm, using the process-based KM model adapted from Burnett et al. (2004, pg. 29; 2013) and further expanded upon with reference to the McKenzie and van Winkelen model (2004).

The model presented by Burnett et al. tends to confirm the findings of Alavi and Leidner (2001, pg. 114) and has been selected as the KM process model because it includes all of the major KM-related processes that were identified in the KM literature review as being necessary for an organization to gain competitive advantages and improve its well-being (topics which are further discussed in Section 3.2.7). The inclusion of the 'application and exploitation' process is a very important part of the overall model. Moreover, the Burnett et al. model clearly shows the connections between each KM process and, in addition to the inclusion of the key application and exploitation process, views the knowledge creation process as the centerpiece of the model. This view is in line with the view adopted in this research that, in addition to the creation of operational/business knowledge, it is also critical to create (and, later, act upon) knowledge regarding relevant business conditions. Such scanning of the business environment and attempting to make sense of it appears to be the key prerequisite for achieving organizational resilience (iJet International Inc., 2008, pg. 5; McCann et al. 2009, pg. 45; Hamel & Valinkangas 2003, pg. 3; Sundstrom & Hollnagel, 2006, pg. 9).

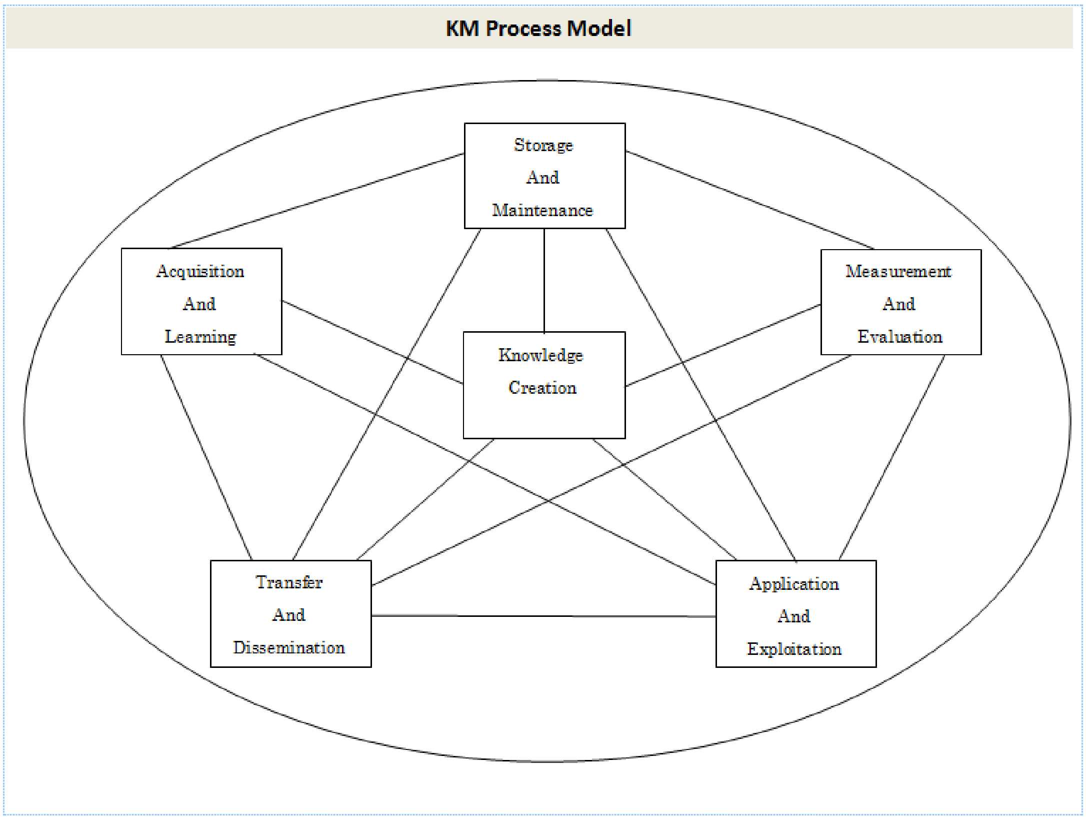


Fig. 3.2.3.1.1: KM processes. [Derived from Burnett et al (2004, pg.29; 2013).]

The expansion to the Burnett et al. (2004) model selected for this research (presented in Appendix IV) comes from the work of McKenzie and van Winkelen (2004). McKenzie and van Winkelen propose a model for leveraging the knowledge resources contained within an organization as well as for the improvement of operational effectiveness within the knowledge economy (knowledge economy as the driver of business growth and productivity leading to overall improved business performance). The process-based model proposed by McKenzie and van Winkelen utilizes six competence areas (namely competing, deciding, learning, connecting, relating and monitoring) that are divided into two categories: those that are internal to an organization (encompassing the first three competence areas) and those that are external to an organization (composed out of the last three competence areas). The uniqueness of the model (which makes it greatly appealing as a viable model for the purpose of this thesis) is the fact that it considers two opposing forces acting on each competence area, which create tension. One force attempts to utilize and maximize the returns from and value of existing knowledge (therefore, it does not abandon the existing goals) and the other force pulls towards change, emphasizing the future and the future value to be derived from knowledge. Competence is attained when both forces act equally and the tension is stabilized. Worth noting is the realization that paying too much attention to either of the force produces a polarized response that is detrimental to an organization (McKenzie & van Winkelen, 2004, pg. 3).

One important point to note is the fact that the 'classic KM-process based models', such as the one adopted from Burnett et al., do not make an explicit distinction between the need to focus on both maximizing the benefits offered by existing KM processes and thinking forward and planning for the future. Robb (2000, pg. 27) emphasizes this point by noting the importance of both planning for the future as well as maximizing the existing opportunities, stating that "[a] resilient organization is able to sustain competitive advantage over time through its capability to do two things simultaneously:

- Deliver excellent performance against current goals, therefore maximizing current opportunities.
- Effectively innovate and adapt to rapid, turbulent changes in markets and technologies, therefore preparing for the future.'

The tension forces present in the McKenzie and van Winkelen model fill this gap, as the forces in all six competence areas establish balance between the current state of things and the state of things to come. Therefore, for the purpose of this research, the Burnett et al. model is used as the base model; however, it is extended by the McKenzie and van Winkelen model in order to provide a mechanism for considering current business goals and future strategic business directions and initiatives.

Similar to the tension forces of McKenzie and van Winkelen, the work of Birkinshaw and Gibson (2004, pg. 47) discusses the concepts of adaptability (as an ability to react quickly to new opportunities) and alignment (creating value from extant organizational capabilities and resources). The attribute combining both adaptability and alignment is referred to as 'ambidexterity' (pg. 47). The concept of ambidexterity is expanded upon in Section 3.3.2. Analogously, Lubatkin et al (2006, pg. 648) view ambidexterity as a composite of exploitation and exploration, similar to McKenzie and van Winkelen's tension forces and Birkinshaw and Gibson's (2004) alignment and adaptability. In addition, they emphasize the KM processes view by relating ambidexterity to the work on Nonaka and Takeuchi (1995) SECI model. According to Lubatkin et al (2006, pg. 648) exploitation involves the use of explicit knowledge bases and their internalization and combination to meet the current needs of existing customers. Exploration involves the use of tacit knowledge bases and their externalization and combination to develop future capabilities and marketing initiatives.

3.2.4 Position of the Existing KM Work in Relation to Technological, Organizational and Ecological Viewpoints

Section 3.2.2 identified numerous schools of thought within the KM field. One of the possible approaches to classifying these viewpoints was based on their main focus, dividing them into three categories: technological (those that considered technology as the driver of the KM field), organizational (those with a focus on the organization in promoting KM) or ecological (those that focus on people, their interactions and the environmental system). As this thesis focuses primarily on the organizational (KM and OR) and technological (DM) aspects, the literature review included an investigation into existing work in order to determine the extent to which technological, organizational and ecological views have received attention in the KM field, as well as to assess to what extent a focus on technology could assist in answering research question #1.

Nonaka and Takeuchi's (1995, pg. 73) spiral model of knowledge creation falls under both the organizational approach and the ecological approach, as their four-phase knowledge creation model (which was introduced in Section 3.2.2) draws on some organizational design aspects and also relies heavily on human interaction.

The work of Malhorta (1998) focuses on the synergies between technology and a business organization's behavioral issues. The author states that technology is a mandatory KM component when it comes to understanding changing business conditions. The importance of environmental scanning, which refers to making organizations aware of changes around them, has been also emphasized by Mallak (1988, pg. 9), Hamel and Valinkangas (2003, pg. 3) and Robb (2000, pg. 30). Robb considers such scanning a necessity for exploring environmental change within the OR context.

Murray (2002, pg. 70) shares this view of technology as an enabler of KM; however, he does not consider it to be the focus of KM, as he emphasizes the role of people: 'performance only improves when people do things differently' (2002, pg. 77). Because the work of Murray focuses on the DIKAR (data, information, knowledge, action, results) and RAKID (reversed order of activities) models, it is primarily based on the ecological view of KM.

McElroy's (2003, pg. 6) knowledge lifecycle model is comprised of a knowledge management layer, a knowledge-processing layer and a business processing layer and appears to be a primarily ecologically based model.

Similarly to that of many other writers, the work of McKenzie and van Winkelen (2004) primarily focuses on people (in the organizational setting), their interactions and the environment.

The literature review reveals that the technological-based view of KM is no longer as prevalent as it once was, as a shift toward the ecological/organizational approach has occurred. These findings tend to be reflected in the recent definitions of KM itself, which focus less on the role of technology and more on organization, people and business strategy (Sundstrom & Hollnagel, 2006, pg. 9; Carlucci & Schiuma, 2006, pg. 36; Wu et al., 2010, pg. 398.)

In their empirical findings, Crook et al. (2011) focus mainly on the ecological aspects of KM. Vatafu (2010), similarly to Crook et al. (2010) and Chou (2011), stresses the importance of intangibles in today's business environment and advocates the non-technological KM focus.

Finally, a number of writers, including McKenzie and van Winkelen (2004), Carlucci and Schiuma (2006), Vorakulpipat and Rezgui (2008), Ibrahim and Reid (2009), Noel (2009), Vatafu (2011) and West and Crook et al. (2011) appear to take the ecological stance by emphasizing the role of KM in value creation through, for example, improved business processes.

The analysis of the above-mentioned authors did not make any direct contributions to answering the first research question, but it did indicate that the technological aspect of KM is perhaps no longer receiving as much attention as it once did. Instead, more emphasis is placed on KM's value-creation role, making this research more important.

3.2.5 KM in Relation to the McKenzie & van Winkelen Framework

Carlucci and Schiuma (2006, pg. 36), based on their literature review, recognize process-based writing about KM (what they refer to as the dynamic view) as being mainstream. So, while work on process–based KM exists, no research was found that attempted to map KM literature onto the McKenzie and van Winkelen model. In addition to the goal of filling this gap, this section attempts to validate the mapping of the KM processes presented in the Burnett et al. (2004; 2013) model onto the McKenzie and van Winkelen model (the mapping is presented in the Appendix IV).

In relation to the six competence areas model used in this research (presented in the previous section), Nonaka's (1991) work very strongly supports the 'competing' area and the first of that area's conflicting goals: the creation of new knowledge. The second pulling factor, the exploitation of existing knowledge, is also supported in Nonaka's work by the illustration of the introduction of a handful of products by an organization that were a market success and that utilized knowledge created in the organization. It can also perhaps be argued that, since knowledge creation is taking place, learning should occur as well. Should such an argument be accepted, then it could be said that Nonaka's work also addresses the 'learning' competence area. Going one step further, one can also expect that conversion from tacit-to-explicit should involve the 'relating' competence area, especially the conflicting goal of paying attention to the close ties allowing for the 'outside-in' and 'inside-out' knowledge flows.

Expanding on Nonaka's work, Frappaolo's (1998, pg. 19) four KM application areas (intermediation, externalization, internalization and cognition) map well onto some of the six competence areas used in this research and include the establishment of mapping onto competing, deciding and connecting.

In terms of mapping the six competence areas onto the framework of Gupta and McDaniel (2002), the 'competing' competence area (knowledge creation) maps well onto the harvesting component; the other component of the 'competing' competence, the exploitation of the existing knowledge, also maps well onto the application component. The relating competence area can also be mapped onto the harvesting component through the collaboration that takes place in harvesting. Finally, the connecting competence area can be mapped onto the dissemination component through the communication channels used by the dissemination component.

In the model presented by McElroy (2003, pg. 6), the comparison between demand-side and supply-side KM is directly reflected in the competence area of the six-competence areas model in which there are opposing forces between the sharing of existing knowledge and the creation of new knowledge.

Within the categories proposed by Schlogl (2005, pg. 3), consisting of Management, Information Sciences, Information Systems and Information Management Classics, one can also see a somewhat limited, but nonetheless possible, mapping of six competence areas. Within the technology-oriented area, composed primarily of the planning, organizing and control of tasks necessary for the provision and usage of IT, the competing and deciding competence areas tend to be the easiest to map. Within the content-oriented category, playing an integrating role between all different aspects, connecting and relating appear to be the most dominant areas of competence. Finally, in the KM category proposed by Schlogl with a main focus on behavioral aspects of information use and the improvement of staff's creativity (2005, pg. 10), competing and learning appear to be the primary areas of competence mapped onto this major category.

A number of writers (Carlucci & Schiuma (2006), Vorakulpipat & Rezgui (2008), Ibrahim & Reid (2009), West & Noel (2009), Vatafu (2011) and Crook et al. (2011)) discuss the impact of KM on competitive advantage and pay attention to the alignment of KM strategies with overall business strategy in order to generate insights into KM performance. These works are examples of supporting the competing and monitoring competence areas.

3.2.6 Measuring the Performance of KM

With the mapping of the KM process-based writings onto the McKenzie and van Winkelen model clearly established by this literature review, prior to seeking to understand the impact of KM on OR, this thesis attempts to understand how one would measure the output of KM initiatives (in terms of positive influence on business, for example). The work of a number of authors was reviewed; the findings are presented in two tables. The first, Table 3.2.6.1, presents the findings of the literature review conducted by Wu et al. (2010) and is provided here for completeness. The second, Table 3.2.6.2, was compiled by the author of this thesis and, similarly to the first table, lists the authors of each work and their approach to measuring the performance of KM. Both tables list the authors in chronological order.

Wu et al. (2010, pg. 398) use return on assets (ROA, which is calculated by dividing a company's net income by total assets, representing how profitable the company is with respect to its total assets) as a KM performance indicator, citing the work of Bierly and Chakrabarti (2004), in which the authors treat ROA as a common measure of business performance and regard it as one of the key ratios for business analysis.

As pointed out by Wu et al. (2010, pg. 398), quoting the work of Tseng (2008), measurement of KM performance became crucial after the realization that KM provides a roadmap for facilitating strategic organizational learning. With reference to the work of Wu (2010, pg. 398), some of the past approaches to the measurement of KM performance that are mentioned in the literature include the following:

Author(s):	KM Performance Measurement Approach:
Bierly and Chakrabarti (1996)	Cluster companies into four groups with different knowledge strategies and state that 'innovator' and 'explorer' groups tend to derive more profit then 'exploiter' and 'loner'.
Choi and Lee (2003)	Look at the non-financial aspects of corporate performance attributed to KM and state that the 'dynamic' KM style results in better performance.
Lee et al. (2005)	Propose the use of a knowledge management performance index (KMPI) for assessing the performance of KM at some point in time, stating that KMPI can represent the efficiency of the knowledge circulation process.
Choi et al. (2008)	Argue for the impact of KM on organizational

	performance suggesting three types of relationships among KM strategies: 'non-complementarity', 'non-critical symmetric complementarity' and 'asymmetric complementarity'.
Law and Ngai (2008)	Examine relationships between knowledge sharing and learning behaviors and their effects on business performance, business processes and on products and service offerings.
Lina and Tsen (2005)	Focus on implementation gaps in the knowledge management system and its impact on corporate performance.
Harlow (2008)	Proposes the tacit knowledge index (TKI) to measure the impact of tacit knowledge on organizational performance, stating that the relationship between higher TKI and financial measures is not very clear.

Table 3.2.6.1: Summary of the literature review conducted by Wu et al. (2010)

In addition to the literature review conducted by Wu et al., the literature reviewed for this research identified the following attempts to measure the performance of KM:

Author(s):	**KM Performance Measurement Approach:**
Sveiby (1997)	Proposes creation of balance sheet for intangible assets of an organization.
Skyrme and Amidon (1998)	Companies seeking to measure the contribution of KM need to focus initially on the value proposition. (Areas of consideration include the market value of information, possible impact of KM on organization [in the case of loss or theft, for example], and potential to increase revenue/reduce costs.)
Hughes and Holbrook (1998)	The objectives of this work were, 1), to develop analytic tools for examining regional systems of

	innovation for policymakers (in Canada) and, 2), to identify and design new indicators of innovation and knowledge creation in this context.
Skyrme (1999)	Discusses the ABBA (assets, benefits, baseline, action) approach to the measurement of intangibles (KM).
Perry and Guthrie (2000)	Measure and report on KM from the cost/benefit perspective. Ask the following question: Within an organization, who is positioned to perform the measurement?
Liebowitz and Suen (2000)	Call for more research on KM metrics. Seek to address 'knowledge level' and the types of value-added knowledge that individuals obtain.
Lee et al. (2005)	Propose a new metric, the knowledge management performance index (KMPI) to evaluate KM at a point in time.
Vestal (2002)	Focuses on measuring KM's effect on business results and less on KM activities.
Marr et al. (2003)	Suggest ways of identifying and evaluating resource transformations in organizations, in order to better understand and manage knowledge creation in order to grow an organization's intellectual capital. Found that the less relevance a person attaches to the KM system, the less the KM system positively impacts the organization.
Kankanhalli and Tan (2004)	Review KM metrics and identify areas where gaps in understanding exist.
Oliveira and Goldoni (2006)	Relate KM metrics to the knowledge management process phases.
Patton (2007)	Looks at extended functionality of balanced scorecards and strategy maps in order to measure

	the performance of KM initiatives.
Marr (2007)	Discusses the preference for the use of indicators rather than ‘hard measures’ for intangibles, including KM.
Ramirez and Steudel (2008)	Propose a simple mathematical model for quantifying knowledge work by calculating a knowledge work score that positions each worker in the knowledge work continuum.
Dolfsma and Leydesdorff (2008)	Use negative entropy, which is a measure suggested by information science for determining the extent to which a system is self-organized.
Andone (2009)	Measures the impact of KM on corporate performance by tying the measurement of KM with the overall corporate performance measurement. Use of balanced scorecards, return on investment and employee surveys.
Chen at al. (2009)	Use an approach that integrates the analytical network process with balanced scorecards from four perspectives (customer, internal business, innovation and learning and financial perspective).
Shannak (2009)	Should measure knowledge management in the same way as any other asset. Need to use performance indicators (performance based on the use of KM strategy) to measure KM.
Handzic (2009)	Aims to improve understanding of the value of KMS from the perspective of individual decision makers involved in time series forecasting.
Kopelko et al. (2009)	Measure firm performance anf the impact of KM on efficiency.
Kulkarni and Freeze (2010)	Developed KM capability assessment (KMCA) instrument based on the 5-level capability maturity model of the Software Engineering Institute. Levels are 1-possible, 2-encouraged, 3-

	enabled/practiced. 4-managed, 5-continous improvement.

Table 3.2.6.2: Summary of KM literature, listing approaches to measuring KM

From the summary presented in the table above, it can be seen that, over the years, numerous approaches have been suggested for measuring the performance of KM. One of the possible (not mutually exclusive) groupings of the presented approaches is the following:

- According to the research of Carlucci and Schiuma (2006, pg. 37), there appear to be three main approaches to linking KM and business performance: the assessment of the likely impact of KM on performance, quantitative measures of the impact of KM on performance and the analysis of causal relations between KM and organizational performance.

Using the groupings proposed by Carlucci and Schiuma, the readings can be classified as follows:

- The work of the following writers can be classified into the group which focuses on the assessment of the likely impact of KM on performance: Bierly and Chakrabarti (1996), Liebowitz and Suen (2000), Choi and Lee (2003), Kankanhalli and Tan (2004), Marr (2007); Shannak (2009) and Handzic (2009).
- The following authors focus on the analysis of the causal relations between KM and organizational performance: Skyrme and Amidon (1998), Skyrme (1999), Vestal (2002), Choi and Lee (2003), Marr et al. (2003), Lina and Tsen (2005), Oliveira and Goldoni (2006), Law and Ngai (2008), Kopelko et al. (2009) and Shannak (2009).
- The quantitative measurement of the impact of KM on performance is examined by the following works: Hughes and Holbrook (1998), Perry and Guthrie (2000), Lee et al. (2002), Kankanhalli and Tan (2004), Patton (2007), Harlow (2008), Dolfsma and Leydesdorff (2008), Ramirez and Steudel (2008), Andone (2009), Chen at al. (2009) and Kulkarni and Freeze (2010).

Somewhat unique, yet probably still classifiable as a form of quantitative measurement, is the approach taken by Sveiby (1996), who attempts to measure the performance of KM performance in an accounting-like method (using the concept of an accounting balance sheet)

Significantly, the review of the literature reveals that there is at present no published research that examines the measurement of the impact of KM on OR, which makes this research a key contributor of such knowledge and therefore a key contributor to the fields of KM and OR.

3.2.7 Role and Value of Processed-based KM Within Organization

The literature review presented in this section was conducted in order to investigate the impact of KM on organizations and the possible value that can be derived from the KM initiatives, with the focus on KM initiatives involving KM processes rather than the other views presented earlier in this chapter. The literature review also serves the purpose of possibly determining (directly or indirectly) the impact of KM on the topic of this research, being OR, which was introduced in Chapter 2 and is discussed in greater detail in Section 3.3. In case of the indirect impact of KM on OR, the review sought the impact of KM on the "compatible with OR" concept; such indirect association is further described in Sections 3.3 and 3.4.

Based on the findings of the literature review, which focused on the role and value of KM within an organization, it is possible to classify the literature into four distinct groups.

The group of writers who focus on the competitive advantages that arise as an end result of KM initiatives is the largest and includes the following writers: Barney (1995), Gupta and McDaniel (2002), Hussain et al. (2004), Anonymous (2006), Carlucci and Schiuma (2006), Fink and Ploder (2007), Vorakulpipat and Rezgui (2008), Ibrahim and Reid (2009), West and Noel (2009), Chou (2011) and Vatafu (2011). While these writers saw the value of KM in allowing organizations to achieve competitive advantages (and all of the associated benefits), for many the road to such an end result varied greatly.

Gupta and McDaniel state that 'knowledge management is a strategic process, which implies the goal of differentiation from competitors such that competitive advantage is forged' (2002, pg. 3). As such, they hypothesize that their five component sequential framework (the authors acknowledge that, in business, things do not necessarily happen in a linear fashion) of activities is essential in effective KM (2002, pg. 3). In their view, the proposed framework leads to better management decisions and organizational activities that ultimately positively affect an organization's net income and market share. Those activities include harvesting (acquiring knowledge from within or from outside an organization), filtering (to exclude unnecessary and irrelevant knowledge), configuration (organizing and storing of knowledge), dissemination of knowledge and application (applying the knowledge to business activities). Gupta and McDaniel's view of the role of KM in generation of the value for organizations through gaining competitive advantages matches the views of the writers identified in the previous paragraph.

Most recently, there has been substantial development regarding the extension of KM, with authors focusing on the use of KM for the purpose of value creation and KM's impact on organizational performance, competitive advantage and efficiency improvement (McKenzie & van Winkelen 2004, Ibrahim & Reid, 2009; Carlucci & Schiuma, 2006; Vorakulpipat & Rezgui, 2008; Vatafu, 2011; West & Noel, 2009; Crook et al, 2011, Gehl 2015). McKenzie & van Winkelen (2004, pg. 16) state that '[t]urning knowledge into value is now regarded as the reason for firms' existence.' In addition, as stated by Carlucci and Schiuma (2006, pg. 43), 'the value of knowledge within an organization is related to its application rather than to its possession.' Vorakulpipat and Rezgui (2008, pg. 283) summarize the 'evolutionary path' of KM by stating that, in order for a firm to be effective, it needs to migrate from a knowledge-sharing (what McElroy calls first-generation) to a knowledge creation culture (second-generation); in addition, it needs to move past that point and create 'sustained organizational and societal values'.

Ibrahim and Reid's 2009 work performed a qualitative research study by questioning six senior managers from the car manufacturing industry in the UK. The authors noted that, for at least one company, there was a link between KM practices and operational benefits, mainly due to improvements of manufacturing processes that were due to existing 'codified knowledge' (Ibrahim & Reid, 2009, pg. 569). Another research participant reported reduced lead times and improved quality due to the application of new knowledge to process improvement and the sharing of the knowledge within the organization. This company identified knowledge creation and sharing as a source of competitive advantage because its cars are designed and delivered to the market more quickly due to the efficiencies achieved as a result. Sharing of the best practices within an organization between world-wide locations has been suggested by yet another company as achieving a reduction of the time involved in business processes as a result of not having to 'reinvent the wheel'. Interestingly, the authors acknowledge the multidimensionality of KM and the problem of understanding the links between KM practices and how KM adds value to organizations –making it the ideal solution for business intelligence (BI) tools. Ibrahim and Reid conclude that 'it can be claimed that KM plays a significant role in adding value in the UK car manufacturing industry' (2009, pg. 573). Their work examines both the causal relations between KM and organizational performance and the specific measurement of the effects thereof.

In the opening paragraph of a 2009 work by West and Noel that investigated the impact of knowledge resources on newly formed (technology) organizations, the authors linked KM to organizational performance, stating that '[a] new venture's strategy – and thus its performance – is based upon the knowledge the firm has about the market, its opportunity in that market, and it's appropriate conduct to take advantage of that opportunity' (2009, pg. 1). Focusing on KM, the authors investigated the relationship between the performance of new ventures and the types of knowledge that are important at the start-up phase, as well as the relationship between the sources of knowledge and the new venture's performance. Interestingly, the authors only found a strong association between networking activity and the knowledge obtained via such activity and the new venture's performance. Knowledge creation and knowledge dissemination thus appear to be key KM processes in the context of a new technology venture's performance. The work of West and Noel illustrates the impact of KM on a newly formed technology firm by assessing the effect of knowledge on organizational performance; in addition, they add a quantitative element by presenting correlations between independent variables (the CEO's knowledge relating to industry relatedness, business relatedness, previous start-up experience, networking frequency and networking information newness), control variables (firm's size and age) and the new venture's performance.

Some authors, such as Carlucci and Schiuma, indicate that 'there is no straightforward link between KM and company's performance but rather a complex relationship' (2006, pg. 35) – making it an ideal problem for analysis by data mining tools (as illustrated in Chapter 6). The knowledge assets value spiral (KAVS) framework proposed by Carlucci and Schiuma (2006, pg. 35) offers a step-by-step process for applying KM initiatives in order to improve a company's performance when KM objectives are linked to performance objectives. Their framework is based on the four-step process that uses the skills of a typical analyst to determine the company's targets and the value of its knowledge asset, defining knowledge asset management processes and assessment of performance improvements based on execution of the prior three steps. While the framework offers a highly practical tool for improving performance, the framework is unlikely to function uniformly and in the same fashion for different organizations, as each organization has its own set of objectives, knowledge asset values and knowledge management processes. The lack of a systematic method of applying the framework (to be treated as a system to be applied at any company) can be seen as the framework's weakness and limits its viability for this research. The 'lack of the same hard measures' of various KM approaches leading to improvements in performance was mentioned as a weakness as early as in 1999 by Armistead (pg. 143). Moreover, the framework of Carlucci and Schiuma resembles, at least in the first two steps, the areas for understanding competences as knowledge presented by Armistead (1999, pg. 148), and it also features the analysis of causal relations between KM and improved organizational performance.

As some authors state, KM is a necessary and determining factor in business success and acquiring competitive advantages (Ibrahim & Reid (2009) and Carlucci & Schiuma (2006)), which is a view shared by the author of this work.

A number of writers consider improved organizational effectiveness and efficiency as the benefits that result from KM initiatives, rather than competitive advantages. These writers include Yli-Renko et al. (2001), Frappaolo (2006) and Fahmi and Vivien (2009).

The practical work of Yli-Rentko et al. involved the administration of a survey to 180 young technology organizations and found a positive relationship between knowledge acquisition (a KM process) and new product development, technological uniqueness and sales cost efficiency. The work of Yli-Rentko et al. is therefore illustrates the actual application of KM for value creation in organizations.

McKenzie and van Winkelen (2004, pg. 2) state that 'giving people better access to available knowledge and helping them use it gives our organization an unrivalled opportunity to improve performance.' To achieve such an improvement, McKenzie and van Winkelen propose a model for leveraging the knowledge resources contained within an organization as well as for improving operational effectiveness within the knowledge economy. The process-based model proposed by McKenzie and van Winkelen is described in Section 3.2.5.

The resulting value of an organization's adaptability and/or sustainability was the focus of the work of Malhorta (1988) and Vorakulpipat and Rezgui (2008).

In Malhorta's (1988) view, the 'old economy', characterized by predictable environments with a focus on the optimization of existing operations and/or processes, no longer suffices, due to on-going shifts in the business environment. The current view of business conditions, what Malhorta refers to as the 'new economy', emphasizes understanding and adjusting to changing business conditions. Knowledge management, according to Malhorta, becomes the vehicle for understanding and adjusting to changing environmental conditions: 'KM is a framework within which the organization views all its processes as knowledge processes. In this view, all business processes involve creation, dissemination, renewal, and application of knowledge toward organizational sustenance and survival' (1998, pg. 1).

Based on the outcomes of their research, Vorakulpipat and Rezgui (2008, pg. 291) state that 'KM has major implications in the learning capability of an organization and its ability to adapt to ever changing and competitive business environment.' Clearly, business adaptation is of particular importance to organizational performance and organizational resilience, as defined in Section 2.4.

While the work of Venzin et al. (1998, pg. 29) was primarily focused on knowledge (a concept introduced in Section 2.2) instead of KM, the link between competitive advantage and underlying knowledge is demonstrated by these authors. Venzin et al. discuss strategizing in the knowledge economy (treating knowledge as a key economic resource), noting that 'knowledge, in one form or another, is of central importance to the development of sustainable competitive advantage of companies.' While one might argue the possibility of achieving a sustainable competitive advantage, most will accept knowledge being the key resource in the knowledge economy that must be properly managed; hence, the need for KM.

Recently, Gehl (2015, pg. 413) examined issues related to knowledge sharing with relation to the person creating knowledge and producing value out of data (the data scientist), the amount of data produced by knowledge sharing and organizational behavior with regard to the data the organization owns and how the data are used in knowledge creation by the data scientist.

Gehl (2015, pg. 414) makes an interesting point about information sharing and the sharing of the knowledge worker (defined by Davenport & Prusak (1998) as the person who takes data and information and converts them into knowledge) stating: 'while knowledge might be easily shared, firms will not share the labor used to mine it.' However, as discussed later, this does not necessarily imply the willingness of the organizations to share their knowledge.

An explanation of why knowledge could be viewed as a valuable commodity is provided by Gehl (2015, pg. 418), in the form of a quote from Davenport and Prusak (1998, pg. 6): 'one of the reasons why we find knowledge valuable is that it is close – and closer than any data or information – to action. Knowledge can and should be evaluated by the decisions or actions to which it leads.' Davenport and Prusak's explanation appears to closely match the hierarchical definition of knowledge (Bergeron, 2003) used in their research and described in Section 2.2.

In terms of KM's role and value in an organization, Gehl (2015, pg. 415) points out that tensions and frictions exist between the KM process of knowledge sharing and the product of the knowledge worker in the context of big data. According to the work of Koopman (2013), which is referenced by Gehl (2015, pg. 419), the knowledge worker appears to be in 'reciprocal and incompatible' tension with knowledge sharing as, per Koopman, '[the] knowledge worker is someone that is hard to share, yet the knowledge is something that cannot exist unless it is shared.' Another factor that adds to these frictions and tensions is the fact cited by Gehl (2015, pg. 421), taken from Husted and Michailova (2002), that 'individuals in firms are inherently hostile to knowledge sharing' and practice so-called 'knowledge hoarding' (the refusal of the knowledge worker to share their knowledge). At the corporate level, Ghel (2015, pg. 425) points out that, if corporate data and knowledge are seen as major corporate assets, they will not be readily shared by the corporations.

Finally, Hussain et al. (2004) provide justification for perceiving KM as having a value creation role, which is in line with the observations of Venzin et al. mentioned above. In their work, the authors make the following statement: 'As the whole world (almost) continues to migrate towards a knowledge-based economy, knowledge management has emerged as a methodology for capturing and managing the intellectual assets of an organization as a key to sustaining competitive advantage.'

As can be seen from the literature review presented in Sections 3.2.6 and 3.2.7, much has been claimed regarding the potential role and value of KM practice, yet no attempts have been made to investigate the impact of KM on OR, and no attempts have been made to use the McKenzie and van Winkelen framework to evaluate the impact of KM on organizations.

3.2.8 Summary

While Section 3.6 provides the conclusions to the literature review, addresses the gaps identified in the literature review and provides the answers to research questions #1 and #2, this section summarizes what has been presented in Chapter 3.2.

Since this research seeks to find a solution to a real-life problem within the business domain, the review of the KM literature has been conducted with an appropriate focus.

The chapter began by examining key developments in the KM field from a historical perspective and moved on to discuss the KM frameworks and KM perspectives applicable to this research; it also introduced the KM models of Burnett et al. (2004; 2013) and McKenzie and van Winkelen (2004), which are used in this research. The chapter also presented the mapping between these models, showing their relationships. The chapter then considered the orientation of KM with respect to technological, organizational and ecological views and with respect to the models of Burnett et al. and McKenzie and van Winkelen. Finally, the chapter closed with an extensive review of the approaches to measuring the performance of KM and discussed the role that KM can play within organizations and the value it. The following chapter addresses the next key component of this research, OR.

3.3. Organizational Resilience

3.3.1 Introduction

The need to deal with business uncertainty and the changes caused by globalization and changes in political conditions and demographics, among other factors, brought about the need for studies that address and resolve the business challenges that arise as a result of these changes. The field of OR is one such area of study. The need for OR is appositely expressed by the following quotation from Darwin, as used by Mallak (1998, pg. 8): 'It is not the strongest species that survive, nor the most intelligent, but the most responsive to change.'

This section reviews published works relating to OR that build on the definition of OR presented in Section 2.4. This section therefore provides the foundation for Section 3.4, the following section, which examines the impact of KM on OR through a DM lens.

The literature review in this chapter begins by examining the development of the OR field, seeking to trace the evolution of the field through various approaches to OR. Thereafter, the role, value and application of OR are investigated, which is followed by a review of works relating to the measurement of OR. Finally, the review closes with a discussion of attempts to measure OR, which provides the foundation for the discussion in the next chapter. A graphical representation of the layout of Section 3.3 is presented below, in Fig. 3.3.1.1.

3.3.2 Development of the OR Field

The definition of OR presented in Section 2.3 focused on the business context and on organizations that perform well under both favorable or adverse business conditions, as opposed to resilience that takes the form of responding to some form of crisis. The definition presented in Section 2.3 provides the foundation for this chapter.

As illustrated by the variety of OR definitions given in Section 2.3, the field of OR becomes fragmented when attempting to identify and define the main OR concepts. This is in line with the results of a study of OR by Benn (2011, pg. 5), who, in addition to noting the fragmentation of the OR field, also acknowledges the relatively recent emergence of the field of OR from the field of organizational theory.

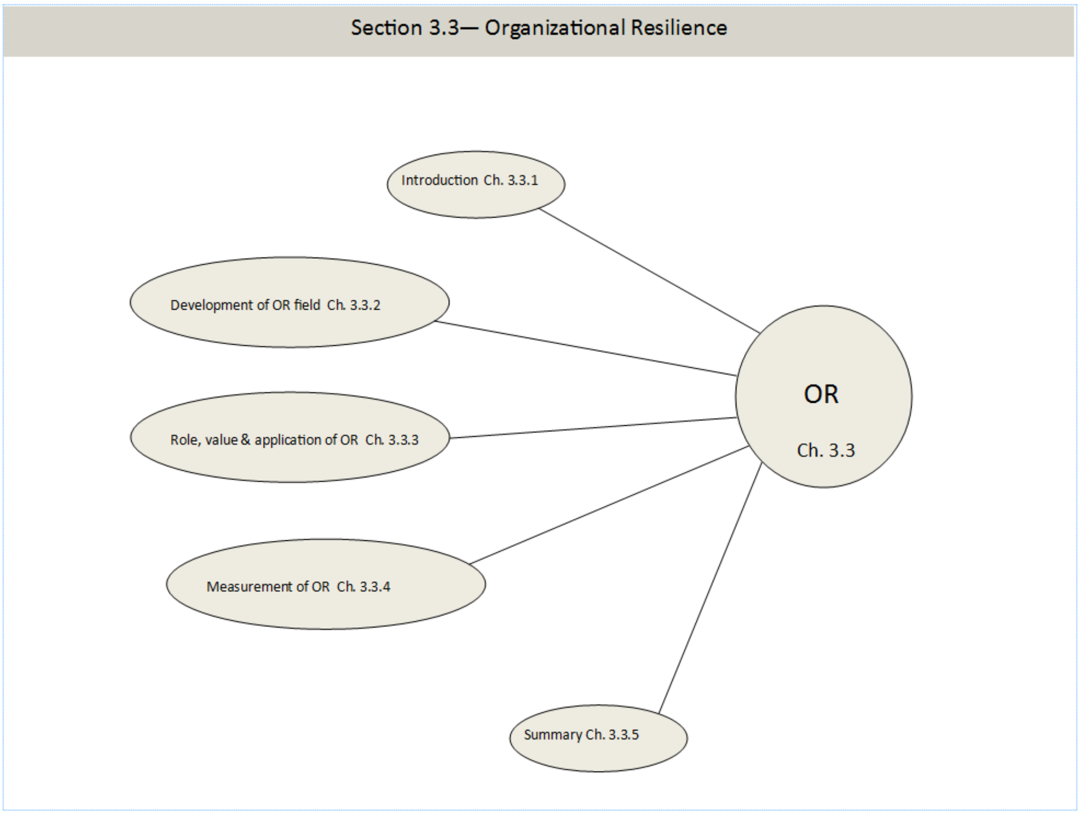

Section 3.3— Organizational Resilience
Introduction Ch. 3.3.1
Development of OR field Ch. 3.3.2
Role, value & application of OR Ch. 3.3.3
Measurement of OR Ch. 3.3.4
Summary Ch. 3.3.5
OR
Ch. 3.3

Fig. 3.3.1.1: Graphical representation of the contents of Section 3.3

In one of the earliest works related to OR, Horne (1997) notes that, in order for a firm to remain competitive in the world of the 'new order/new economy,' there is a need to change the business's operation from a resource-based and/or optimization-focused approach to business operations to a more balanced one that promotes resilience as well as productive capacity and optimization. Horne (1997, pg. 26) states that, when a firm is viewed as a system, '[p]roductive capacity will continue to be important to organizations, but it must now take place in a much more balanced order of things. Becoming a 'learning organization' has much to do about learning about your own system's resilience.' Given Horne's (1997, pg. 27) definition of OR, as presented in Chapter 2.3, and its reliance on the detection and dissemination of environmental change and its emphasis on the concept of the learning organization, Horne's work might very well have been the first attempt to link KM to the field of OR.

Horne (1997, pg. 27) contributed to the study of OR by providing a definition of OR, context for the study of OR and the introduction of the OR 'common strands' that aid in the development of a sustaining framework, with most of the strands directly mapping onto the six competence areas that form the framework of this research. (The six competence areas are discussed in greater detail in Section 3.2.3.) In summary, Horne (1997, pg. 27) states that, due to the uniqueness of each organization (and its systems), there is no simple one-fits-all formula for developing resilience, but, rather, '[r]elationships within an organization and how information flows along these relational paths is a key element in the development of resilience.' Again, there appears to be an indirect reference to the 'transfer and dissemination' process of Burnett et al. (2003, 2013) that can be mapped using the mapping presented in Fig. 3.2.7.2 onto McKenzie and van Winkelen's six competence model's competing, learning and connecting areas.

Similarly to Horne, Mallak (1998, pg.9), in his roughly contemporary work, refers to two main forms of organizations: organic and mechanistic. The mechanistic organization is characterized as 'a machine': 'efficient, programmed, with low level of uncertainty in a closed system design'. The organic organization resembles a living organism: 'complex response, flexible, and higher levels of uncertainty [are found] in an open system design'. So, in the environment of high uncertainty and change, the organic organization appears to be better suited for survival.

In his work, Mallak (1998) presents and argues for resilience principles intended to be used for implementing OR in an organization. The principles he offers are based both on reviews of resilience literature and practice and include the following: perceiving experiences constructively, positive adaptive behaviors, the provision of adequate resources, expanded decision-making boundaries, the practice of bricolage, the development of a tolerance for uncertainty and the building of a virtual role system. Many of these principles, such as adaptive behaviors and expanded decision-making boundaries, can be mapped onto the competing and deciding competence areas of McKenzie and van Winkelen's (2004) KM model.

Interestingly, Mallak emphasizes the importance of resilience as a force and/or method for dealing with uncertainty and change, yet he does not place an emphasis on environmental scanning as a necessary component leading to OR. Moreover, in his critique of the existing literature (Mallak, 1988, pg. 9), he states that, in the literature of the field, there should be less time devoted to environmental assessments and more to developing resilient organizations and individuals. In addition to the lack of importance placed on environmental scanning, Mallak's work does not make any connection between the KM and OR; however, it does offer important OR principles that can be applied to today's organizations.

The work of Robb follows that of Horne, Orr and Mallak and takes a more balanced perspective of OR. Robb presents a framework that is based on two components/systems: the performance system, which is responsible for performance of current goals and tasks associated with day-to-day operations, and the adaptation system, which is responsible for the long-term survival of an organization (Robb, 2000, pg. 27). His balanced approach arises from the fact that he asserts that both the adaptation and the performance systems are needed in order for an organization to be resilient.

Robb (2000, pg. 27), taking a somewhat more systematic point of view than previous writers, realizes the importance of both planning for the future and maximizing existing opportunities. He states that the 'resilient organization is able to sustain competitive advantage over time through its capability to do two things simultaneously:

- Deliver excellent performance against current goals, therefore maximizing current opportunities.
- Effectively innovate and adapt to rapid, turbulent changes in markets and technologies, therefore preparing for the future.'

Robb's discussion of the tension between performance skills and adaptation skills, the two complementary sets of fundamental skills that a resilient organization should actively develop (Robb, 2000, pg. 30) is a concept that is, to some extent, reflected in the framework used in this research. This research's framework, which is based on the work of McKenzie and van Winkelen (2004, pg. 6) and presented in Fig. 3.2.7.2, also uses the concept of tension, referred to by the authors of the framework as 'conflicting pulls'. The conflicting pulls used within in this project framework occur in all six competence areas. As stated by McKenzie and van Winkelen (2004, pg.3): 'Generally, one aspect of the tension pulls towards stability and the delivery of current value from knowledge; the second largely supports change and the creation of future potential value from knowledge.' Moreover, many of the skills that can be taken from Robb's concept of tension can be directly mapped onto one or more of the competence areas identified by McKenzie and van Winkelen.

The concept of ambidexterity introduced in Section 3.2.3.1 provides additional insights into what McKenzie and van Winkelen refer to as 'conflicting pulls' (2004, pg. 6). The two components that make up ambidexterity, alignment and adaptability, are responsible for exploiting values (or reducing costs) from current organizational resources and moving towards new opportunities (Birkinshaw and Gibson, 2004, pg. 47). While the work by McKenzie and van Winkelen applies the conflicting pulls to the six competence areas, the work of Birkinshaw and Gibson introduce the additional composition of ambidexterity: structural ambidexterity (different organizational structures for different activities/products) and contextual ambidexterity (choosing between alignment and adaptation orientated activities). In relation to the work of McKenzie and van Winkelen (2004) as well as the work of Robb (2000), Birkinshaw and Gibson emphasize both individual employee and the entire organization as a source of ambidexterity, and view the structural and contextual separations as complementary. Similarly, Lubatkin et al. (2006) use the concept of exploitation and exploration as analogous to the alignment and adaptability to show how top management team's (TMT) behavioral integration through ambidextrous orientation positively affects organization performance (measured by growth in sales, growth in market share, return on equity and return on assets).

As highlighted by Lubatkin et al. (2006, pgs. 648, 652), the main criticisms with regard to ambidexterity relate to the view that attaining and maintaining proper balance between exploitation and exploration is not an easy task, and that the pursuit of ambidexterity does not guarantee subsequent performance. The findings of Lubatkin et al. (2006, pg. 666) suggest however that TMT's behavioral integration is the key in achieving an ambidextrous orientation in SMEs that leads to the improved OP.

From the systematic view point adopted in the work of Sundstrom and Hollnagel (2006, pg. 9), in which resilience is an attribute or property of a system, 'the property of resilience implies that a system has the ability to maintain a healthy state over time despite the fact that it (or these wholes) may be subjected to negative and/or destructive events'. (By 'wholes,' the authors mean organized entities.) This concept of a healthy state, perhaps represented by a balanced system, differs significantly from the balancing of the tensions in the McKenzie and van Winkelen (2004, pg. 3) model. Sundstrom and Hollnagel's model has more to do with system based re-balancing as opposed to McKenzie and van Winkelen's concept of competence, in which the tensions between the need to maintain the stability of business systems and the drive for change are in balance.

The work of Hamel and Valinkangas (2003) focuses mostly on change within a business environment and the constant need for businesses to 'make their future' by aligning their strategies to constantly changing opportunities and trends. Hamel and Valinkangas (2003, pg. 3) point out that 'any organization that hopes to become resilient must address four challenges:

- The cognitive challenge – a company must not be too attached to its past as well as to be humble so that it can properly interpret and react to the changing business environment.
- The strategic challenge – a company needs to be aware of the changes around it and needs alternatives to follow in response to such changes.

- The political challenge – a company needs to be able to divert resources from yesterday's products and services to tomorrows.
- The ideological challenge – a company needs to look beyond the operational excellence and flawless execution.'

The theme of Hamel and Valinkangas' work is the importance of the ability to think beyond current business operations and optimization focus. In addition, the authors stress the need for environmental scanning and adjustment to environmental changes, which had been part of the work of Horne (1997), Mallak (1998), Robb (2000), McKenzie and van Winkelen (2004) and appears to be one of the key aspects of OR.

The work of Starr et al. comes from the practitioner's perspective, as it was written by the senior 'risk management' members of Booz Allen Hamilton, Inc. In their paper, the authors discuss enterprise resilience (ER) and systematic resilience (SR), where ER is the ability and capacity to withstand systematic discontinuities and adapt to new risk environments (Starr et all., 2003, pg. 3), while systematic resilience is the ability to understand an organization's interdependencies and to foresee and plan around the discontinuities that can occur within them (Starr et al., 2003, pg. 5). In their discussion, the authors focus on a discussion of resilience in the context of risk, mainly as a disruption to the primary earning drivers. Similarly to other 'resilience definitions,' they emphasize the need to align organizational strategy, operations, management systems, governance structure and decision-support capabilities so that risks can be detected (Starr et al., 2003, pg. 3). The novel aspect of their point of view comes from the fact that 'traditionally, risks have not been perceived in the context of key earning drivers, but rather in broad categories, each of which was managed in functionally isolated way' (Starr et al., 2003, pg. 4). Their view allows for the integration of the 'risks' managed by the CIO, CFO and COO along with looking at interdependencies between risks spanning multiple functions in the organization that also affect OR. While representing a unique OR context, the work of Starr et al. does not provide any direct or indirect links to KM or the models used in this research.

A contribution to the field of OR also came from outside the business field. A paper by Friedman (2005), who, at the time of writing, was a clinical and corporate psychologist, uses 'human factors' as the lens through which to view OR. Friedman argues that 'an organization can only be resilient if its human capital is resilient and that the features of resilient organization include:

- Powerful, flexible innovative leadership.
- Sustainable internal alignment (mainly through open communication).
- Capacity for leadership and workforce to accept the challenges, roll with the punches and bounce back.'

Interestingly, the three features necessary for resilience presented by Friedman have a number of similarities with the work of other authors. Horne (1997, pg. 27) discussed the strands that are required in order for the organization to be resilient. Horne's strands, similar to Friedman's features, include communication, coordination, commitment and connections. Mallak (1998, pg. 10), when listing his 'resilience principles', discussed the need for positive adaptive behaviors and for practicing bricolage. Robb, on the other hand, discussed visioning, the exploration of environmental change and its implications, creativity, experimentation and inquiry (2000, pg. 30). Finally, Hamel and Valinkangas pointed out the need for organizations to address challenges in order to become resilient. Most of Friedman's features could perhaps be classified as strategic challenges and could be best mapped to McKenzie and van Winkelen's (2004, pg. 31) six competence model, primarily the competing competence – through flexible, innovative leadership and the workforce being willing to accept the challenge.

There is, however, one speculative aspect of Friedman's paper. The quotation above, 'an organization can only be resilient if its human capital is resilient,' is a possible point of disagreement, especially when taking into account Coutu's (2002, pg. 52) point of view that '[v]alues (from the value system of a firm), positive or negative, are actually more important for organizational resilience than having resilient people on the payroll. If resilient employees are all interpreting reality in different ways, their decisions and actions may well conflict, calling into doubt the survival of their organization. And as the weakness of an organization becomes apparent, highly resilient individuals are more likely to jettison the organization that to imperil their own survival.'

The work of Sundstrom and Hollnagel builds on von Bertalanffy's general system theory, with resilience being a non-directly observable property of a system (Sundstrom & Hollnagel, 2006, pg. 4). The work of Senge is also very important to their work, particularly Senge's concept of system thinking, which the authors quote: 'seeing interrelationships rather than linear cause-effect chains, and processes of change rather than snapshots' (Sundstrom & Hollnagel, 2006, pg. 9). They also use the concept of feedback loops, which can be described as a circular view of cause and effect of actions, feeding upon each other and forming a circular pattern of behavior. Sundstrom and Hollnagel provide their view on how companies can learn to facilitate the development of resilience. The authors also draw attention to the implications of adopting a systematic approach to organizations and/or business systems. They begin by referring to the resilience of organisms as the 'highest form of resilience,' which they state is a property that organic systems have. They formulate the analogy of a business system as an open (organic) system based on its need to exchange information and/or resources with the external environment. The authors also stress the importance of the control component within a business system for the purpose of monitoring all points of contact with the external environment. The emphasis that Sundstrom and Hollnagel place on environmental scanning and on checks and balances aligns well with the views of Hamel and Valinkangas (2003). They also map directly onto the monitoring competence of the McKenzie and van Winkelen (2004) model.

The report published by iJet (2008, pg. 5) provides interesting insight into the evolution of resilience. The paper discusses resilience in terms of its evolution, including the following phases, from least to most desired: reactive, proactive and adaptive. The lowest level on the path to an organization becoming resilient is the 'disaster recovery' type of response (also known as disaster response) where the primary purpose is to respond and recover, and there is little concern for the continuation of operations. The next form of action on the way to becoming resilient is the proactive form (which refers to business continuity). Here, companies focus on continuing operations and the preservation of revenue. The final form of action, which can truly be considered a form of resilience, is the adaptive form (or business resiliency). This form, the 'actual resilience,' concentrates on revenue preservation and on the pursuit of business opportunities. While iJet presents an interesting paper regarding the evolution of resilience, the paper does not explore the role of KM in any of the three evolution phases.

Recently, Braes and Brooks proposed a project that would make significant contributions to the field of OR, as it intends to identify the essential concepts that contribute to making an organization resilient as well as the essential concepts that form the philosophy of OR. In short, the authors plan to 'organize' the main concepts utilized in the field of OR due to the fragmentation with the field. The view of Braes and Brooks regarding the fragmentation of work within OR field is shared by the author of this thesis, as the following claim from the literature appears to still hold: 'There is little consistency in its use in terms of organizational resilience and a lack of common understanding as to the essential concepts prevails' (Braes & Brooks, 2010, pg. 15).

Some recent work, including that of Ponis and Koronis (2012), extends the concept of resilience beyond the consideration of a single organization to consider the resilience of an entire supply chain. In their paper, Ponis and Koronis set out to conceptualize supply chain resilience (SCRes) and identify which supply chain capabilities can contain disruptions and how these capabilities affect SCRes. The direction of current research towards understanding the interconnectedness of organizations and their impact on a single organization as well as a whole industry is not surprising, given the trends in recent years of minimizing inventories (cost efficiency) and operating in a 'just-in-time' (JIT) fashion. Clearly, the JIT movement has some advantages, but it also carries with it many risks. Interestingly, in 2003, Starr et al. (2002, pg. 5) had already discussed 'interdependence risk,' defining it as 'unanticipated risk exposure across the extended enterprise that is beyond an individual organization's control. Examples of interdependence risks include supply chain disruptions, government interventions, and public infrastructure destruction.' Later in their work, Starr et al. coined the term 'systemic resilience,' referring to a firm's 'ability to understand its interdependence and plan around discontinuities that can occur within them' (Starr et al., 2003, pg. 5). As this thesis focuses on the resilience of a single organization (or, more specifically, on the impact of KM on OR), the concept of SCRes may appear beyond the scope of this work. The resilience view of SCRes as a method used to prevent some undesired event differs vastly from the OR concept studied, which is the ability of an organization to remain in business (and perhaps even flourish) under adverse business conditions. Despite the divergence of SCRes and this study's focus on OR, the concept of SCRes must be addressed, as it reflects the state of contemporary research in the general area of resilience; it also draws attention to the fact that, in today's interconnected world, a company might fail if its supply chain, or part of it, fails. The case of Erickson and Nokia brought up by Ponis and Koronis (2012) serves as an example of such a failure.

As previously discussed (and viewed through the lens of OP within this research) the concept of ambidexterity has emerged as a new research paradigm in organizational theory (Raisch et al, 2009, pg. 685), leading to a rapid increase in the volume of related research over the last twenty years (Tran, 2015, pg. 31). Yet, there are a number of controversial issues in regards to the tensions associated with the ambidexterity. Raisch et al. (2009) point out the following tensions with ambidextrous organizations: tension of differentiation (distinct business units for exploitation and exploration) vs. integration (within the same business unit); individual vs. organizational level; static (cycle the focus of activities between exploitation and exploration) vs. dynamic (engage in exploitation and exploration activities at the same time) perspective of ambidexterity; and internal (internal to the organization's knowledge processes) vs. external (external to the organization's knowledge processes) perspective. The concept of ambidexterity and the tensions mentioned above position themselves clearly in relation to the work of McKenzie and van Winkelen (2004) in that the competing competence area can be thought to be analogous, from the KM process perspective, to ambidexterity and the remaining five competencies areas to be key factors in resolving tensions as stated by Raisch et al. (2009).

To understand the ever-increasing organizational tensions created by the competing demands placed on organizations, the paradox lens has been recently introduced by Smith and Lewis (2011, pg. 381): 'Paradox studies adopt an alternative approach to tensions, exploring how organizations can attend to competing demands simultaneously', with the paradox defined by Smith and Lewis (2011, pg. 382) as '[a]s contradictory yet interrelated elements that exist simultaneously and persist over time.'

Finally (and significant to research that involves both applied research and well defined business issues) another topic which has recently gained significant attention in academic writing, is the issue of dynamic capabilities of organizations. As stated by Teece and Leih (2016, pg. 7), 'Dynamic capabilities enable the firm to integrate, build, and reconfigure internal and external resources to address and shape rapidly changing business environments.' Such capabilities are about doing the right things versus doing the things right (Teece and Leih, 2016, pg. 7) and are manifested by organizations that are built to respond to the unexpected, what Teece and Leih refer to as hallmarks of strong dynamic capability.

Because the field of OR tends to draw from a number of different domains (such as engineering, economics and psychology), there are different approaches to grouping views of OR. One such grouping is the classification of writings based on the most common domains that each work references to a significant extent (which is not necessarily the same as the predominant domain from which the work originates). To illustrate various methods/approaches for achieving OR, the following table is presented:

Domain:	**Writers:**
Psychology	Mallak (1998), Coutu (2002), Friedman (2005)
Biology	Sundstrom & Hollnagel (2006)
Engineering/System view	Horne (1997), Horne & Orr (1998) , Mallak (1998), Robb (2000), Sundstrom & Hollnagel (2006)
Risk management	Starr et al.(2003)
Business/Economics	Robb (2000), Hamel & Valinkangas (2003), McDargh (2003), Starr et al (2003)., Birkinshaw and Gibson (2004), iJet (2008), McCann et al. (2009)
Multidisciplinary	Braes & Brooks (2010), Benn (2011), Cockram & van Den Heuvel (2012)

Table 3.3.2.1: Summary of OR authors by domain

In addition to the categorization given above, one can classify work based on its emphasis on a given OR component or that most frequently found in the literature:

Emphasized Component/Aspect of OR:	**Writers:**
Individual	Horne (1997), Mallak (1998), Horne & Orr (1998), McDargh (2003), Birkinshaw and Gibson (2004), Friedman (2005), McCann et al. (2009), Braes & Brooks (2010), Cockram & van Den Heuvel (2012)
Organization	Horne (1997), Horne & Orr (1998), Coutu (2002), Hamel & Vailnkangas (2003), McDargh (2003), Birkinshaw and Gibson (2004), Sundstrom & Hollnagel (2006), iJet (2008), McCann et al. (2009), Braes & Brooks (2010), Cockram & van Den Heuvel (2012)
Enterprise/Supply chain	Starr et al. (2003), McCann et al (2009)., Ponis & Koronis (2012)
Culture/Structure	Horn & Orr (1998), Robb (2000), Hamel & Valinkangas (2003), Birkinshaw and Gibson (2004), McCann et al. (2009), Braes & Brooks (2010), Cockram & van Den Heuvel (2012)

Table 3.3.2.2: Summary of OR authors by emphasized OR component

While the above-presented classifications are informative, one other aspect discussed in some OR literature is important: paying attention to the existing business and current business conditions. An organization cannot simply disregard the business that is currently operating and its environment and simply focus on anticipating the future and making plans for it. In the reviewed literature, few of the writers considered both the need for a business to satisfy current goals as well as the need to prepare and plan for the future. Of the group of OR-related authors whose work was examined, Mallak (1998), Robb (2000), McCann et al. (2009), and, to lesser extent, Hamel and Valinkangas (2003) are the writers who consider both aspects in their discussions of OR. This aspect neatly fits into the six competence model of the KM writers McKenzie and van Winkelen (2004), in that the consideration of current business goals as well as making preparations for the future corresponds to McKenzie and van Winkelen's concept of tensions between exploiting existing knowledge in the competence area (to optimize current business goals) and creating new knowledge (to meet future business goals).

From the above review of OR-related literature, it can be stated that there is no one-method-fits-all approach for achieving OR. The methodologies and approaches summarized in the tables above also differ in several areas. For this research project, the most appropriate approach to OR appears to be the combined view of Robb (2000) and Hamel and Valinkangas (2003), as, when combined, they represent the view of OR that is the most appropriate for this work. The OR model offered by Robb considers both maintaining current operations in the best possible manner (the performance system) as well as generating new options for the organization (the adaptation system). In addition, Robb's model also relies on skills and organizational culture as the foundations for the performance and adaptation systems; this is in line with the personal views of the author, as, without either the right skills or the right culture, very little will be achieved in terms of OR. The view of Hamel and Valinkanagas contributes to a complete understanding of what it means and takes for an organization to be resilient. They (implicitly) expand the concept of organizational culture by specifying four challenges (cognitive, strategic, political, and ideological) that need to be overcome in order for an organization to become resilient. Other extremely important OR elements presented by Hamel and Vailnkangas are the concepts of environmental scanning (in order to detect change) and the need for variety and alternatives in response to the findings of such environmental scanning. Clearly, firms need to know how various environmental changes can affect them and must have options to respond to such changes, which goes back to Darwin's quote from the beginning of this chapter. The OR lens selected for this research relies on views of Robb (2000) and Hamel and Valinkangas (2004), as their views are well aligned with the KM model of McKenzie and van Winkelen (2004) and the views of the author of this thesis; these authors' concepts are a critical to this work due to its emphasis on the aspects of OR identified by these writers, and this is reflected in the questions of the questionnaire used in this research.

3.3.3 Role, Value and Application of OR

This section focuses on the investigating the role of OR in organizations, the value derived or to be derived from the OR and the application of OR within an organizational setting. The findings discussed in this section inform this work's pragmatic approach to the deriving of value from OR and position the discussion of the OR component of this research. The findings also form the foundations for further discussion in Section 3.4, which considers the impact of KM on OR as seen through a DM lens. Finally, the content of this section provides the guidance in the formulationof the OR-related questions used in this research questionnaire.

According to Horne (1997, pg. 27), '[t]o varying degrees, resilience is a fundamental quality found in individuals, groups, organizations, and systems as a whole. It allows a positive response to significant change that disrupts the expected pattern of events without resulting in regressive/nonproductive behavior'.

The work of Mallak (1998) appears to be similar to that of Horne (1997) in that, in its description of OR, it also emphasizes the need for the adaptive positive capabilities that are needed in order for an organization to remain competitive.

Hamel and Valinkangas (2003, pg. 13) note that technological discontinuities, regulatory upheavals, geopolitical shocks, industry deverticalization and disintermediation, abrupt shifts in consumer tastes and hordes of non-traditional competitors are the factors that compel companies to frequently reinvent.

Of interest, and somewhat unique in the OR literature, is the reference of Hamel and Valinkangas (2003, pg. 7) to variety (strategic alternatives) as a key component of resilience. They state that 'resilience depends on variety'. As an analogy, they use the variety of life forms as a mechanism for the survival (resilience) of life on the planet despite the many adverse conditions that existed and events that occurred in the past. In addition, companies must guard against strategy decay by being replicated, supplanted, exhausted and/or eviscerated (Hamel & Valinkangas, 2003, pg. 7). The role of OR as a response to and/or defense mechanism against environmental changes is emphasized in the four challenges presented by the authors, along with the value derived from the strategic alternatives, which is seen by the authors as a key OR component.

The following quotation from Hamel and Valinkangas (2003, pg. 13) directly relate to the areas of competence presented by McKanzie and van Winkelen: 'Any company that can make sense of its environment, generate strategic options, and realign its resources faster than its rivals will enjoy a decisive advantage. This is the essence of resilience. And it will prove to be the ultimate competitive advantage in the age of turbulence – when companies are being challenged to change more profoundly, and more rapidly, then ever before'. In particular, it appears that, in order for a company to be able to enjoy a decisive advantage, the company needs to be competent in all 'six areas of competence': competing, deciding, learning, connecting, relating and monitoring. (This is discussed further in Section 3.4, which maps KM on OR.)

The key takeaway from the work of Sundstrom and Hollnagel, (2006, pg. 9) is the need for an organization to consider resilience in the system context with a system control component, looking at various interdependencies rather than at linear cause and effect and seeing the entire process of change rather than snapshots. Finally, the authors' call for environmental scanning as well as for checks and balances tend to align with the views of Hamel and Valinkangas (2003) presented earlier.

A more recent justification for the importance of OR is presented by McCann et al. (2009, pg. 45): 'We believe that organizations are now seeking greater resiliency because they are overexposed to the environmental turbulence in the form of more frequent and intense competitive and operational disruptions,' where 'environmental turbulence' is defined as '[t]he pace and disruptiveness of change within an operational, competitive or larger contextual environment' (McCann et al., 2009, pg. 45). Despite the definition originating from a relatively recent source, it shares a common theme of 'environmental change' with the definitions already encountered. It can be seen from the work of Horne (1997), Mallak (1998), Hamel and Valinkangas (2003) and McCann et al. (2009) presented in this section, as well as in the previous section, that the role of OR is to successfully address uncertainty that arises due to the changing business environment and to provide methods for organizations to remain resilient.

Robb (2000, pg. 27) also realizes that his proposed framework represents an ideal state towards which organizations seeking to be resilient should work. In addition to presenting his framework, Robb (2000, pg.29) sees culture, skills and the architecture of each of the two systems as the integrating components between them; they are also elements necessary for achieving OR.

A lack of agility, argue McCann et al., (2009, pg. 45), can result in organizations operating more slowly and less productively. In addition, the authors point out the impact of knowledge on organizational performance, stating that the inability to retain top talent with critical skills can have a highly negative effect on organizational performance. The 'newness' in the work of McCann et al. also comes from their perspective on resilience: They do not focus on resiliency at the individual level; rather, they consider multiple levels (individual, team, organization and industry). While McCann et al., write with a focus on OR, they do identify the positive effects of knowledge, and therefore KM, on OR through positive impact on organizational performance.

The practitioner paper written by iJet (2008) presents actions taken by resilient organizations that appear to map very well onto the McKanzie and van Winkelen (2004) framework selected for this research (this framework is discussed in Section 3.2.3 and in Chapter 4). The topics discussed by iJet include (iJet, 2008, pg. 5), and map onto the framework, as follows:

- Using predictive intelligence for early warnings and situational awareness. This can be mapped onto the monitoring and learning competence area of the McKenzie and van Winkelen (2004) model used in this research;
- Resilient organizations desire a common operating platform across various business entities for a global perspective on risk and opportunities. This can be mapped onto the connecting competence area of the model used in this research;
- Resilient organizations routinely communicate with their stakeholders. This can also be mapped onto the connecting area of the McKenzie and van Winkelen model, as well as the relating competence area; and
- Resilient organizations' actions stretch beyond response and recovery and seek to identify business opportunities that disruptions may

represent. This can also be mapped onto the deciding competence area of the model used in this research.

According to the iJet definition of resilience presented in Section 2.4.1, the role of resilience is to provide organizations with an adaptive ability; for the iJet authors, this is more meaningful than merely responding to and recovering from environmental disruption.

The extract from the work of Braes and Brooks (2010) highlights the need for resilience, particularly during extremely adverse business conditions, such as those that existed in the United States of America after the 2008 financial collapse caused by subprime mortgages. In their discussion, the writers address the events of 2008 that deeply affected the United States' economy and markets and their impact on organizations. The significance of Braes and Brooks work is the fact that this study intentionally selects the companies that were in existence during that time frame, or immediately after, and asks questions about these organizations' performance and/or actions during those challenging business times. The work of Braes and Brooks, therefore, validates the choice of questions that focus on determining the impact of the financial crisis of 2008 on organizations in this work's questionnaire. Braes and Brooks, commenting on the events of 2008 and the markets' consequent loss of $US17 trillion in value, state '[t]hese types of events have highlighted the need for organizations to become more innovative or adaptive in their attitude to proactive strategies, thus ensuring more effective prevention, enhanced protection, increased preparedness, effective mitigation, increased response capacity and streamlined recovery process; is short organizations, need to become resilient' (2010, pg. 17).

3.3.4 Measurement of OR

The purpose of this section is to review the literature for any practical insights into how one would actually go about the measurement of OR; any insights discovered can be used in the development of the OR-related questionnaire questions used in this research. This review of methods of measuring of OR is also conducted in order to validate the methodology chosen for this research (as discussed in Chapter 4).

One of the earliest examples of applied research in the area of OR in organizational settings was the work of John Horne and John Orr (1998). The underlying premise of their early OR-related work was a 'system-based' view of an organization as a living system and the people within the organization as the elements capable to respond to major change, which could function as a measure of the effectiveness of their organization. One of the outcomes of their work was the 1996 74-item organizational resilience inventory assessment tool, which was designed to identify the occurrence of behaviors associated with system resilience in organizations (1998, pg. 34). As a result of their work, they proposed seven streams of resilient behavior that contribute to the development of resilience in an organization (1998, pg. 31): community, competence, connections, commitment, communication, coordination and consideration.

The work of Horne and Orr, especially their 74-item assessment tool (1998, pg. 34), has some resemblance to this research project. The authors' tool was designed to identify the occurrence of behaviors associated with system resilience in organizations. In addition, the tool evaluates the level of importance of each of these streams of resilience to the overall system and also attempts to identify regressive behaviors. The successful use of such a tool by Horne and Orr validates the choice of the research instrument used in this thesis; this research makes use of an 84-item tool that is based on McKenzie and van Winkelen's (2004, pg. 6) model. The model is comprised of six competence areas, three internal and three external. With regard to this grouping aspect, there is a similarity in methodology between Horne and Orr's work and this project in that the questions used are based on grouping the actions and/or events that are thought to be contributors to and/or enablers of OR. However, the main difference is the fact that this project attempts to determine an organization's 'level' of OR by asking questions, based on the literature review, that address the 'OR enablers' in an organization. That is, the assumption, based on the literature review, is that, if an organization possesses and/or conducts most of the elements and/or actions identified by the researchers as being necessary to achieve OR, it should be fairly resilient as a result. Finally, there is also a similarity in the fact that both projects set out to identify the enabling and/or regressive attributes and/or behaviors that impact OR.

Despite the fact that their work is a propositional study, the paper of Braes and Brooks (2010) provides substantial material to address their first objective, the identification of the concepts that contribute to OR, when they discuss perspectives on OR from various domains and various locations around the world. They attempt to identify common ground between these perspectives; for example, they attempt to show the correlation between resilient people and resilient organizations. The authors have also started work on their second objective, the identification of the essential concepts that form the philosophy of OR. The challenge that they identify is similar to that encountered in addressing the first objective of their work: '[the] essential concepts of OR are not clearly understood' (2010, pg. 18). To accomplish their second objective, the authors propose the use of grounded theory four-phased method that involves a review of current standards related to OR, interviews with OR resilience experts, a survey of OR-related practitioners and a comparative study of earlier phases. As a starting point for their work in relation to the second objective, the authors propose four sets of characteristics and/or essential concepts of OR, which are grouped into the following categories: organizational (interdependencies or situational awareness), contributors (the fields contributing to characteristics, such as emergency management and enterprise risk and management); tactical (such as risk identification, risk avoidance and emergency response) and strategic (leadership, communication or culture and values). The expected outcomes of the study proposed by Braes and Brooks are stated as follows (2010, pg. 20): 'The outcomes of the proposed study are expected to be an authority's summarization of OR, delivering a comprehensive set of essential concepts that must be present to make an organization resilient.'

Another, more recent, practical work in the area of OR is that of McCann et al. (2009). In this work, the writers report the results of a study of 471 North American companies. Based on their extensive research, the authors demonstrate that 'environmental turbulence may indeed be managed by building agility and resiliency. Companies exhibiting higher levels of agility and resiliency are more competitive and profitable, even with higher levels of turbulence' (McCann et al., 2009, pg. 45). (The authors define agility as the capacity for moving quickly, flexibly and decisively in anticipating, initiating and taking advantage of opportunities and avoiding any negative consequences of change.)

The work of McCann et al. has a profound influence on this research, especially in validating the approach and methodology selected for this work. The work of McCann et al. has many similarities to this work, primarily in the research methodology adopted and the topics researched. The authors set out to measure agility and resilience and their relation to organizational performance and how various levels of turbulence affected such relationships. This work attempts to identify how the impact of KM on OR might be measured using DM, and, while there is no intention of explicitly accounting for agility, one might find that measuring instruments that focus on promptness of actions (agility) are used at multiple points in this thesis. However, the studies addressed in this work's research into indirect and unintentional measurement of what McCann et al. make no provision for measuring outcomes based on different levels of disturbances. The length of the questionnaires used in both studies also differs, as McCann et al. ask 30 questions, whereas this research uses 84 compiled questions, but both instruments ask questions of a similar audience of senior executives and/or key decision-makers. In their research, McCann et al. used two organizational performance measures, competitiveness and profitability, as dependent variables. In this research, organizational resilience is the dependent variable (measured using concepts such as profitability) and the six competence areas are the independent variables. The time frame in both studies appears to cover the period of extraordinary financial crisis in the US that began in 2008. Finally, similarly to the work of McCann et al. that investigated impact on operational agility, this research sets out to determine various relationships between KM practices and OR and their strengths; however, this research, for the reasons stated in Chapter 4, uses DM as the primary analytics tool.

The results obtained by McCann et al. tend to support the works presented in the last section, including the OR-related texts of Horne (1997), Mallak (1998), Hamel and Valinkangas (2003), and McCann et al. (2009), as well as the KM-based writers McKenzie and van Winkelen (2004), which primarily state that companies that are sensitive to their business environments and respond to environmental changes (which, if they are considered agile organizations, they will be able to do quickly and decisively) are more competitive and profitable.

Given the findings, it would be intriguing to see what results the authors would be able to derive through the use of the DM as tool for analysis. The findings of this research could assist in such analyses.

3.3.5 Summary

This chapter examined another component of the overall range of topics that inform this research, namely OR. Similarly to the prior chapters, the discussion about OR focused on the business context and began by considering developments in the OR field. Thereafter, the role, value and application of OR within an organizational setting were discussed. The chapter closed with a review of the literature that discussed attempts to measure OR; this forms the foundations for the Section 3.4, which examines the impact of KM on OR through a DM lens.

3.4 Data Mining and its Impact on KM and OR

3.4.1 Introduction

While there appears to be a limited number of academic and practitioner publications that deal with data mining, knowledge management and organizational resilience at the same time, as well as a limited number of texts regarding the use of data mining tools for the purpose of analyzing the impact of KM on OR, the following literature review attempts to present the 'current state of academic and professional literature' in these areas.

Specifically, this literature review builds on the introduction to DM provided in Section 2.5 and attempts to answer research questions #1 and #2 by addressing the following issues:

- The application of data mining with respect to KM and/or OR;
- The feasibility of using data mining to assess the relationship between and/or impact of KM or OR, seeking to identify insights into the aims of this thesis;
- What prior work exists regarding above two issues from both the theoretical and practical perspectives?; and
- What techniques can be used to measure OR as well as the impact of KM on OR?

The structure of this chapter, which focuses on reviewing the literature regarding the use of DM in a business setting, is presented in Fig. 3.4.1.1. Given its very well-defined objectives, this chapter's contents are presented in a more structured manner than preceding chapters, with the findings presented in a slightly different ways: sections acts as category-holders for published works. This section is primarily composed of two main sub-sections, preceded by a general discussion of the DM/BI field. The first sub-section examines the theory-based utilization of DM with respect to KM and OR. The second part of Section 3.4 attempts to investigate the practical aspects of the utilization of DM with respect to KM and OR. That is, considering that KM is an independent variable and OR the dependent variable in this study, this section seeks to develop an understanding not only of the use of DM to measure the impact of KM on OR but also to measure the impact of DM on KM (an independent variable) as well as to measure the impact of DM on OR (the dependent variable). During the discussion, references to the argument established in this section of this work (Section 3.4), equating OR to organizational performance and competitive advantage, are made where appropriate. Sub-section 3.4.3 examines the key factors of this study from the theoretical perspective, while sub-section 3.4.4 investigates the same factors from the applied perspective; both sub-sections provide guidance for this research. Section 3.4 ends with an extended summary of the works analyzed, while Section 3.6 summarizes the entire literature review presented in Chapter 3.

Note that, as pointed out in Section 2.5.2, the terms BI, BI&A, PA, and DM are used interchangeably throughout this research.

The graphical organization of Section 3.4 is presented in Fig. 3.4.1.1, below.

3.4.2 Development of the DM Field in the Context of this Research

The literature review performed by Shollo and Galliers (2013, pg. 2) indicates that, up until the time of the review was conducted, there were two main parallel perspectives on how researchers viewed the role of BI in organizations and its impact. The first perspective involved perceiving BI from what Shollo and Galliers (2013, pg. 2) refer to as the 'traditional view', which views BI

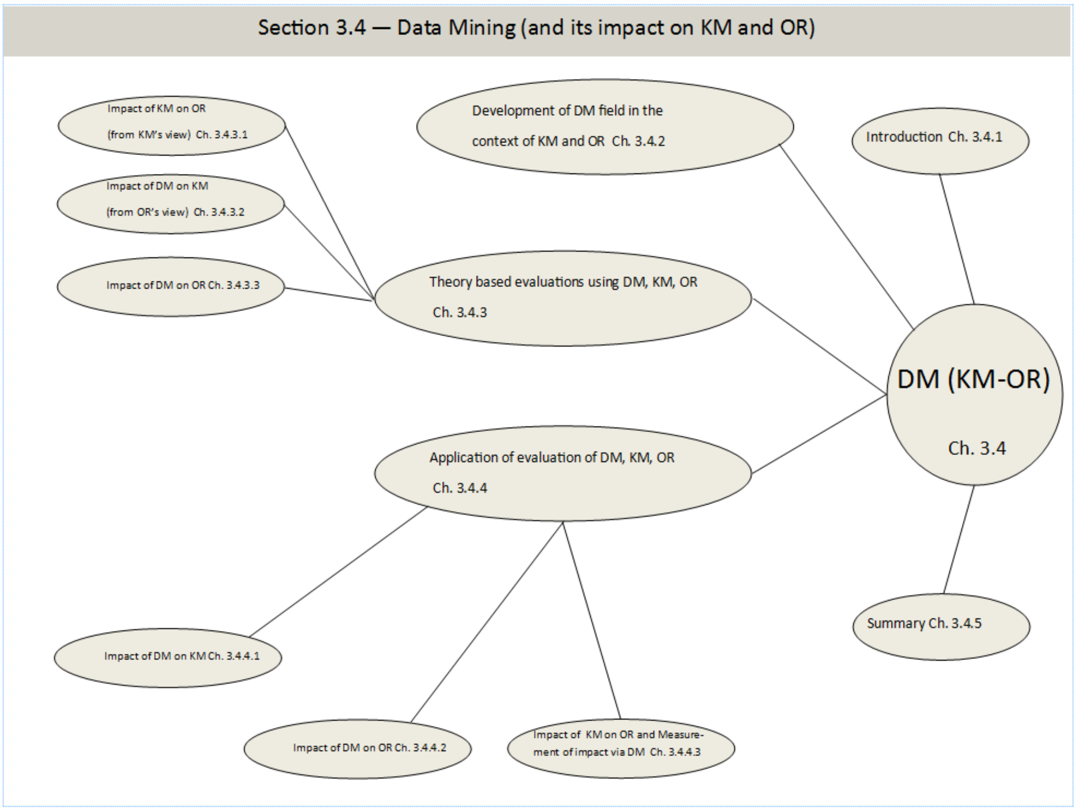

Section 3.4 — Data Mining (and its impact on KM and OR)
Impact of KM on OR (from KM's view) Ch. 3.4.3.1
Development of DM field in the context of KM and OR Ch. 3.4.2
Introduction Ch. 3.4.1
Impact of DM on KM (from OR's view) Ch. 3.4.3.2
Impact of DM on OR Ch. 3.4.3.3
Theory based evaluations using DM, KM, OR Ch. 3.4.3
DM (KM-OR) Ch. 3.4
Application of evaluation of DM, KM, OR Ch. 3.4.4
Summary Ch. 3.4.5
Impact of DM on KM Ch. 3.4.4.1
Impact of DM on OR Ch. 3.4.4.2
Impact of KM on OR and Measurement of impact via DM Ch. 3.4.4.3

Fig. 3.4.1.1: Organization of Section 3.4

systems as decision-making enhancers due to BI's role in the transformation of raw data into information and the transformation of information into knowledge (this agrees well with Bergeron's [2003] hierarchical knowledge model, which is used in this research and was presented in Chapter 2.2.1). This view presents technology as a catalyst for various DM technologies, techniques and data sources that contribute to BI.

The second perspective, according to Shollo and Galliers (2013, pg. 2), emphasized the importance of people and the process of knowledge creation, referred to as organizational knowing. Because the traditional view stores data and draws attention to the data in a context-free manner that only becomes informative after information retrieval and with the addition of personal knowledge, Shollo and Galliers (2013, pg. 3) state that the traditional technology-based view may facilitate the transformation of data into knowledge but certainly does not enable it, raising questions about the usefulness of knowledge management systems. Second, the human-sense-making perspective, what Shollo and Galliers (2013, pg. 3) refer to as knowing, stresses the importance of BI systems as facilitators in knowledge creation and learning processes, as these processes are social and participative in nature. Shollo and Galliers (2013, pg. 2) provide a brief history of the evolution of ICT towards BI, presented in Table 3.4.2.1 below:

Development Era:	**Management support systems:**	**Purpose:**
Mid 1960s	Management information systems	Provided structured, periodic reports and information to support structured decisions.
Late 1960s	Decision support systems	Provided decision-related information to support semi-structured or unstructured decisions.
Early 1970s	Model-based DSS	Optimization and simulation models to improve managerial decision-making.
Late 1970s	Document-based systems	Enabled the searching of documents to support decision-making.
Late 1970s	Executive information Systems	Provided predefined information screens for senior executives.
Early 1990s	Data warehouse Systems	Provided large collection of historical data in organizational repositories enabling analysis.
Early 1990s – 2000s	Knowledge management systems	Managing knowledge in organizations for supporting the creation, capture, storage and dissemination of information.
2000 – Present day	Business intelligence systems/Business analytics	Decision support linked to analysis of large collections of data based on integration of different systems and data sources.

Table 3.4.2.1: Historical perspective of BI. [Derived from Shollo & Galliers (2013, pg. 3).]

In the present day, business analytics, listed in the last row of Table 3.4.2.1 above, is further divided into three main areas: descriptive, predictive and, most recently, prescriptive analytics. These three ideas are described by Evans and Linder (2012) as follows:

Management Support Systems:	**Purpose/ Illustrative References:**
BI & A: Descriptive form	Summarizes data into meaningful charts and reports. Use the data to understand past and current business performance. Typical questions descriptive analytics help to answer: How much did we sell in each region? What was our revenue last month? In the BI&A industry, descriptive analytics is typically thought of as a method for describing what has happened.
BI & A: Predictive form	Analyzes past performance in order to predict the future by examining historical data, detecting relationships or patterns in the data and then extrapolating these relationships forward in time. Typical questions predictive analytics help to answer: What will happen if demand falls by 10%? What do we expect to pay for milk? In the BI&A industry, predictive analytics is typically thought of as a method for forecasting what will happen.
BI & A: Prescriptive form	Uses optimization to identify the best alternatives to minimize or maximize an objective. Typical questions include: What is the best pricing for an advertising strategy? What is the best mix in a retirement portfolio? In the BI&A industry, prescriptive analytics is typically thought of as a method by which to ask 'how can we make it happen'?

Table 3.4.2.2: Categories of present day analysis

While the results of this research are capable of supporting all three types of analysis listed in Table 3.4.2.2, its main focus is on the predictive and prescriptive forms. These forms, in form of model results, are described further in Chapter 6.

The literature review undertaken in this chapter and the discussion in the rest of this section attempts to group the readings thematically into the following groups, which are not mutually exclusive: works directly related to DM and KM; works directly related to KM and organizational performance as well as competitive advantage, per the argument stated in the next section linking these concepts to OR; and works that indirectly relate DM to KM or OR. The focus of Section 3.4.3 is on the theoretical nature of the literature addressed, while the following Section 3.4.4 focuses on the applied aspect of DM. Note that, because of the gap in the literature regarding the impact of KM on OR when viewed through the DM lens, this category is of minimal length and forms the contents of Section 3.4.4.3. This research, therefore, directly addresses such a shortage of articles.

To establish the context for this chapter's discussion, the definition of BI presented in Section 2.5 and used in this chapter, as well as that used by Isik et al. (2013, pg. 13), is restated here. The definition of BI provided by Watson (2009, pg. 6) as 'a broad category of applications, technologies, and processes for gathering, storing, accessing, analyzing data to help business users make better decisions' also includes DM, because DM, in its most general form and within a business context, enables and facilitates decision-making. Watson's definition also includes all of the preparatory steps dealing with data loading and data cleaning (which are addressed in Chapter 5).

3.4.3 Theory-based Evaluations of DM, KM and OR

This section examines the theoretical work that addresses the relationships between the key fields addressed in this research. The goal of this section is to provide a sound theoretical backing for the use of DM as an analytical tool when evaluating the impact of KM on OR and for accepting OR as a concept analogous to OP and/or OR. (The argument made in this research is that, if one accepts that OP [organizational performance, efficiency, effectiveness and competitive advantage] = OR, than the impact of KM on OR should be the same as the impact of KM on OP.)

3.4.3.1 Impact of KM on OR

This section provides an initial look into the literature that focuses on the impact of KM on OR but does so without emphasizing the technological (DM) element. In analytical terms, this section reviews the works related to the impact of the independent variable (KM) on the dependent variable (OR).

In Section 3.2, while discussing KM, the foundation was laid for the analogical inductive argument that, since KM impacts various aspects of organizational performance (aspects such as organizational effectiveness, efficiency and competitive advantage, among others) and since OR (as defined in Section 2.4) shares many attributes of the impacted aspects of organizational performance, it can therefore be expected that KM will similarly impact OR.

To establish such an argument, the number of shared features known from the academic research on the impact of KM (on organizational performance, efficiency, effectiveness and competitive advantage) and on OR has been established. The OR work of Horne (1997), Mallak (1998), Robb (2000), Hamel and Valinkangas (2003), Starr (2003) and iJet (2008) has been successfully contrasted with the findings on the impact of KM on organizations derived from the work of Venzin et al. (1998), Armistead (1999), Yli-Renko et al. (2001), Gupta and McDaniel (2002), Hussain et al. (2004), McKenzie and van Winkelen (2004), Anonymous (2006), Frappaolo (2006), Fink and Ploder (2007), Ibrahim and Reid (2009), Vatafu (2011), Crook et al. (2011) and Chou (2011).

From the work of authors listed above, one can expect that a direct link exists between KM and certain aspects of organizational performance. In particular, from the literature review, it can be stated that KM, when successfully implemented and properly managed, improves organizational efficiency and effectiveness. It improves adaptation to changing business conditions and leads to competitive advantage.

Because there is a gap in the literature that deals with the impact of KM on OR, it is necessary to refer to alternative methods of establishing the relationship between KM and OR. One such tool is analogical inductive argument. To complete the inductive argument that KM impacts OR, it must be shown that OR resembles many aspects of organizational performance (OP).

Then, since OP and OR are similar, if not identical, then it is expected that KM impacts OR in the same way it impacts OP.

When discussing OR, Mallak (1998, pg. 8) states that resilient organizations implement effective actions to advance. In addition, resilient organizations implement positive adaptive behaviors quickly in order to adapt to the immediate situation. While Mallak's view of OR possibly maps onto several aspects of organizational performance, the strongest match for Mallak's view of OR and the aspects of organizational performance is found in the work of Vorakulpipat and Razgui. It can be said that, in Mallak's view, OR leads to an organization adapting to changing business conditions. Similarly, Horne (1997, pg. 27) stresses the need for resilient organizations to be able to detect environmental changes quickly and to employ adaptive responses early.

Green (2006, pg. 267), based on a literature review, acknowledges the importance of knowledge management for achieving organizational benefits, mainly in the areas of improving performance and competitive advantage, and proposes a conceptual model of the knowledge valuation system.

One of the challenges identified by Green (2006, pg. 276) is the fact that the retrieval of information from repositories and making sense of the retrieved data need to occur within the domain context and with the intended use in mind. This observation agrees with the DDDM introduced in Section 2.5.6, and it also conforms well to the consideration for 'constraints' presented by Cao and Zhan (2006).

Hamel and Valinkangas (2003), Starr et al. (2003) and iJet (2008) align OR with the view of KM's impact on an organization held by Vorakulpipat and Rezgui (2008) and Chou (2011) by stressing the importance of business adaptation to the constantly changing business environment and the issues that arise as a result.

Having examined the similarities between the results of KM's impact on an organization and OR, it can be stated that there are many attributes that are shared between them. Moreover, because the impact of KM on an organization has been successfully established by other writers, it is possible to expect KM to impact OR in the same fashion it impacts OP.

Finally, when computed, the relationship between organizational performances (labeled as such as a group of questions in the questionnaire used in this research) and OR has been measured for this study to be 0.763, implying correlation. (Illustration of correlation can be found in Appendix II and Fig. A3.35 in Appendix III.)

Wu et al. (2010, pg. 397) cite Bierly and Chakrabarti (1996), Choi and Lee (2003) and Lee et al. (2005) and state that past attempts to measure the performance of KM and the relationship between KM styles and corporate performance was done using traditional statistical methods. The study also points out that there are only a few works that utilized the DM approach, although the references for these works are not provided (Wu, 2010, pg. 401). The purpose of this research is to fill this void.

The review of the impact of KM on OR can be also presented from what the author of this research sees as the OR writers' viewpoint.

Lee (2008, pg. 111), acknowledges the positive effect of KM on organizational performance stating that, when KM is effective in an organization it enhances products, speeds product deployment, improves operational efficiency, increases sales as well as profits and improves customer satisfaction. Lee's work (2008, pg. 111) also proposes a KM architecture and discusses how combined DSS and DM can greatly enhances KM.

According to Lee (2008, pg. 124), when business goals are aligned with knowledge processes (particularly the processes of knowledge creation, structuring, disseminating, and application), organizations can grow strategically.

The work of Lee (2008, pg. 113) proposes an architecture for such enhancements to KM. With the help of the proposed framework, the role of DW and DM will not only be the facilitation and codification of knowledge but also the great enhancement of the retrieval and sharing of knowledge within the enterprise (Lee, 2008, pg. 132). An important component is the part of the framework containing the feedback loop in the knowledge warehouse component, which allows for enhancement of the knowledge stored there as the function of passed time and for the tested and approved knowledge to be fed back.

The application of the new theories is also visible in the literature related to the key aspects of this research. A paper by Choi et al. (2008) examines the relationship between KM strategies and organizational performance using a novel approach: complementarity theory, which is taken from economics. Other new theories are also introduced and are presented in the applicable sections that follow. The novel concept presented by the writers of strategic knowledge management (SKM) represents a unique combination of concepts and constructs from the strategy, knowledge management, information systems and data-mining literature (2012, pg. 67).

While the strategic knowledge model presented by Moayer & Gardner (2012, pg. 72) contains no details about the DM algorithms that could be used in the model, the DM component of the model Moayer & Gardner plays a primary role in that it is utilized in the iterative learning process. The iterative learning process, when integrated with strategic knowledge management (and informed by the market, shareholder, resource and knowledge based views), can lead to competitive advantages. What is interesting, and in line with the observations of the author of this thesis, is that the model considers four views as inputs to strategic KM: the market-based view, the stakeholder-based view, the resource-based view and the knowledge-based view. Considering these various views as part of the DM model (in the form of model dimensions and attributes, for example) greatly enhances the model and will lead to superior DM results that will affect or create competitive advantage.

Perhaps the most important contribution made by the work of Moayer and Gardner (2012) with relation to this research is their concept of the strategic knowledge management (SKM) framework, which incorporates DM with the KM strategy.

The SKM model of Moayer & Gardner (2012) was used as the basis for arriving at the theoretical OR model due to the completeness of its approach. The resulting OR model is described in greater detail in Section 3.5.

When discussing the creation of value by an organization, Gehl (2015) frequently emphasizes a forgotten but key individual in the knowledge creation process: the data scientist.

In his article, Gehl (2015, pg. 413) examines the issues related to knowledge sharing with relation to the person who creates knowledge and produces value from data (the data scientist), the amount of data produced by knowledge sharing and organizational behavior with regard to the data it owns and how that is used in knowledge creation by the data scientist.

Building on the work of Davenport and Prusak (1998), Gehl (2015, pg. 414) makes an interesting point about the knowledge worker (defined by Davenport & Prusak [1998] as the person who takes data and information and converts them into knowledge) that produces value from data: '[W]hile knowledge might be easily shared, firms will not share the labor used to mine it.' This, however, as discussed later, does not necessarily imply that organizations are willing to share their knowledge.

An explanation of the view of knowledge as a valuable commodity is provided by Gehl (2015, pg. 418), in form of quote from Davenport and Prusak (1998, pg. 6): '[O]ne of the reasons why we find knowledge valuable is that is close – and closer than any data or information – to action. Knowledge can and should be evaluated by the decisions or actions to which it leads.'

Gehl (2015, pg. 415) points out the existence of tensions and frictions between the KM process of knowledge sharing and the output of the knowledge worker in the context of big data. According to the work of Koopman (2013), which is referenced by Gehl (2015, pg. 419), the knowledge worker appears to be in 'reciprocal and incompatible' tension with sharing as, per Koopman, '[a] knowledge worker is someone that is hard to share, yet the knowledge is something that cannot exist unless it is shared.' Another factor that adds to these frictions and tensions is the fact cited by Gehl (2015, pg. 421) from Husted and Michailova (2002) that 'individuals in firms are inherently hostile to knowledge sharing', and practice so-called 'knowledge hoarding' (the refusal of knowledge workers to share their knowledge). At the corporate level, Ghel (2015, pg. 425) notes that, if corporate data and knowledge are seen as major corporate assets, they will not be readily shared by corporations. Seeing data and knowledge as corporate assets can have a profoundly negative effect on what McKenzie and van Winkelen (2004 pgs. 148; 155) call 'outside-in' and 'inside-out' knowledge flows of the connecting competence. This negative effect can find expression in terms of comparing practices, participation in intelligence networks, driving common standards and co-operative competition. One of the goals of this research is the ability to capture/measure such negative effects as well as to determine the extent to which they impact on OR.

3.4.3.2 Impact of DM on KM

Referring once more to analytical terminology, this section examines the literature related to the impact of DM (the measuring instrument and the knowledge-creating mechanism, among other aspects) on KM.

The work of Cao and Zhang (2006) focuses on using data mining to generate practical knowledge that is usable and actionable for businesses, as opposed to seeing DM as a purely data-driven methodology that analyzes business issues in an isolated trial-and-error manner. To accomplish the goal of extracting business usable knowledge using DM techniques, Cao and Zhang (2006, pg. 50) present the domain-driven in-depth pattern discovery (DDID-PD) framework.

While the DDID-PD model has some elements of the CRISP-DM model discussed in Section 5.1 and resembles the DDDM framework introduced in Section 2.5.6, it enhances the CRISP-DM model through addressing the comprehensive constraints that impact the studied problem, employing domain knowledge/experts and emphasizing the human role in a process. One can argue that the processes presented by the DDID-PD model of constraint analysis, actionability enhancements, human-mining interaction and the focus of the extracted knowledge on business needs are also addressed by the CRISP-DM model if the work at each step of CRISP-DM is performed diligently. Where the DDID-PD differs from the CRISP-DM model is in what Cao and Zhang (2006, pg. 50) refer to as in-depth modeling. In-depth modeling is, according to the writers, an additional round of data mining. Such a framework could have been considered in this research were it focused on numerical findings, but, from the personal experience of the author of this work, if the industry standard CRISP-DM model is followed diligently, the model is usually sufficient for generating results that are satisfactory both for this research and for commercial purposes.

While an introduction to the similarities and differences between BI and DM was provided in Section 2.5.2, the focus of this section is not on such similarities or differences but rather on the impact of DM on KM. In order to make DM more relevant to BI, a paper by Wang and Wang (2008) investigates the relationships between DM, BI and KM and proposes a knowledge sharing model for knowledge workers.

Wang and Wang's (2008, pg. 623) contrasting of KM and BI is worthwhile, as they state that KM differs from BI in several aspects, the main difference being that KM is concerned with human subjective knowledge rather than data or objective knowledge. In addition to pointing out the key difference between KM and BI, Wang and Wang (2008, pg. 623) also see DM as the bonding agent between the fields of KM and BI. In particular, they state that DM as a BI tool is responsible for knowledge discovery and such discovery is a KM process, as it involves human knowledge (primarily knowledge sharing in the case of DM/BI bonding).

The following statement by Wang and Wang (2008, pg. 624) appears to be the subject of argument: 'DM is considered to be useful for business decision making especially when the problem is well defined'. Brusilovski and Brusilovski (2008, pg. 1), however, state that the best use of DM and the biggest returns come from applying DM against unstructured (meaning not well-defined) business problems. The view of Brusilovski and Brusilovski is shared by other writers, such as Lee (2008, pg. 112), Moayer and Gardner (2012, pg. 69) and Lamont (2015-B, pg. 8), as well as by the author of this thesis.

Additionally, Wang and Wang (2008, pg. 625) recognize the fact that DM and the knowledge generated by DM has its limitations, primarily related to DM's creation of hard-to-apply knowledge and the neglect of the role of business insiders in developing and applying knowledge across an organization.

Wang and Wang (2008, pg. 631) conclude that, for DM to truly become useable as a business organization knowledge discovery tool, it must be integrated with KM, which is now being attempted with the DDDM and DDID-PD DM frameworks (Cao & Zhang 2006, Zhang et al. 2010).

Shollo and Galliers (2013, pg. 1) state that, over the last twenty years, BI and the concepts associated with it have gained significant prominence. The authors also state that 'recent studies provide evidence of increased organizational productivity as a result of BI systems use' (2013, pg. 1). It is unclear, however, what the authors consider as 'BI,' as the term 'BI' has been defined in many different ways, ranging from the 'regular reporting of historical data" to 'prescriptive analytics" (defined in Table 3.4.2.2). The work of Shollo and Galliers (2013) investigates how BI systems affect the process of knowledge creation, or what the authors refer to as 'knowing,' in organizational contexts. The work of Shollo and Galliers (2013), therefore, sets out to investigate knowing in an organizational setting and the facilitating role BI plays. (Shollo & Galliers [2013, pg. 4] define knowing as 'an active process of making new distinctions accepted in organizational settings and embodied in organizational changes, from which learning occurs'.)

In an article, Hopkins and Schadler (2015, pg. 10) discuss the issues associated with the prevailing BI problems and offer some insights regarding how to current situation might be changed so that firms can become insight-driven, relying on systems of insights. According to Hopkins and Schadler, the three main issues are as follows:

- Too much data and too few insights;
- Poor linkage between insights discovered and business action; and
- Scarce learnings from actions taken.

To circumvent these issues, Hopkins and Schadler present their company's (Forrester) 'system of insight' – a combination of business ideas, actions and technology that allows insights to be consistently transformed into action.

The system of insights resembles some aspects of DDDM (Cao & Zhang 2006), in that it emphasizes in-depth analysis but is not as focused on the use of experts in order to arrive at insights that are useful to a business. On the contrary, the system of insights does appear to have 'trust' in technology-based solutions, as its suggestions for improvement focus on improving elements such as infrastructure, data supply and access. One of the authors' last points is that '[m]achine learning and cognitive computing make insights easier to find, test and implement' (Hopkins & Schadler, 2015, pg. 22). The difficulty of making sense from the knowledge generated from DM has been mentioned by many writers recently (Cao & Zhang (2006), Brusilovski & Brusilovski (2008), Adejuwon & Mosavi (2010), Wu et al. (2010), Li et al. (2012), Shollo & Galliers

(2013), Corte-Real et al. (2014), Hopken (2014) and Rao (2015)) and appears to be one of the key current issues in the DM field.

With regards to the DSS, which Lee (2008, pg. 112) defines as 'interactive computer-based systems that help decision makers utilize data and models to solve unstructured problems', Lee states that DSS can also help in the conversion from a tacit to explicit form of knowledge by the creation and investigation of sets of 'what-if' scenarios. Data mining, according to Lee (2008, pg. 112), is a part of DSS, functioning as a decision support tool, the goal of which is the generation of information that is actionable for decision-making from the data stored in data warehouses (DW). Citing the work of Lau et al. (2004), Lee (2008, pg. 113) states that DSS and DM can enhance some of the KM processes. Those processes include tacit-to-explicit knowledge conversion, leveraging of explicit knowledge and explicit knowledge conversion.

Similarly to the work of Choi et al. (2008) discussed in the previous section, the work of Li et al. (2012) involves the application of a new theory.

The work of Li et al. (2012, pg. 2480) cites and utilizes the work of Yang and Cai (2007), which involves the concept of extenics: a method of systematically collecting information and knowledge as well as a series of methods that allow researchers and analysts to utilize collected information and knowledge. The extension theory combines element theory, extension methodology and extensions engineering to address contradiction problems and the formulation of models. The extenics method can be used for the collection of as much data as possible from various sources, reflecting different views or understandings of the problem that DM is used to solve. Such collection of data is referred by Li et al. as knowledge seeding (2012, pg. 2480).

Li et al. (2012, pg. 2483) state that DM can discover new knowledge from databases using what they refer to as seed knowledge (the primary relative knowledge).

An important contribution to this work, as well as to the practical application of DM in the field, comes from the emphasis that Li at al. (2012, pg. 2483) place on the need for the post-processing of data mining results, so-called knowledge cultivating (the process of finding knowledge from the seed or primary, knowledge). The vast amount of data and knowledge available as seed knowledge and generated as the outcome of DM results, according to Li et al. (2012, pg. 2483), in many methods for both knowledge cultivation and for DM, with the most practical method being the one that finds relationships between the information, knowledge and the expected goal; this results in the formation of knowledge trees

Shollo and Galliers (2013, pg. 9) state that BI serves as a balancer of objectivity and subjectivity, given that subjective insights and tacit knowledge are articulated in a way as the backing up of BI results make it more acceptable and appreciated. Moreover, individuals using BI systems derive knowledge from, and make sense of, the data through the processes of data selection, articulation and community (organizational) sense-making.

3.4.3.3 Impact of DM on OR

This section reviews the literature related to the theoretical application of DM with respect to OR (the dependent variable).

The importance of the practitioner-based work of Brusilovski and Brusilovski (2008) for this work lies in their support of DM and DM application stated as:

- The authors support the positioning of the data mining technologies as enablers that allow competitive advantage to be acquired. In support of DM, the authors make the following statement: 'By investing in data mining applications an organization can gain a competitive advantage and uncover information that cannot be identified in any other way' (2008, pg. 1).

  However, the authors realize that data mining software in itself cannot be a source of competitive advantage, as it can be obtained by competitors. Instead, there are certain characteristics that distinguish data mining applications that can lead to this potential competitive advantage. These characteristics include the following:

    - Uniqueness – in the way that the data mining application is used and results analyzed and interpreted;

- Each data miner (the person working with the data mining software) is a unique individual;
- Need for a multidisciplinary team – people from various functional areas are assigned to the DM project. The knowledge that these people possess differs from company to company;
- Synergy of data mining methods – the combination of traditional quantitative approaches with the knowledge discovery offered by algorithms that discover previously unknown patterns in the data is, like the previous factors, unique for a given organization;
- Software dependency – different software vendors implement data mining differently, leading to differences in the output of the algorithms; and
- Role of creativity – the solution to the DM problem leads to non-uniform solutions among data miners.

Ngai et al. (2009) recognize the fact that, over the years, organizations have compiled large amounts of data that are not used for their benefit. Ngai et al. (2009, pg. 2593) quote Berson et al. (2000): 'However, the inability to discover valuable information hidden in the data prevents organizations from transforming these data into valuable and useful knowledge.' Ngai et al. (2009) see data mining as a tool that can be used to help discover knowledge that could be utilized for the benefit of the organization, which matches the views of Hopkins and Schadler (2015).

Adejuwon and Mosavi (2010, pg. 41) appear to conform to the view in the literature that DM can be a source of competitive advantage, as they state '[b]usinesses that can efficiently transform data into useful information can use them to make quicker and more effective decisions and thus form better actionable business strategies which will give them a competitive edge.' Moreover, Adejuwon and Mosavi (2010, pg. 42) see DM 'as a tool of business intelligence by providing the means to transform data into useful and actionable knowledge.'

The work of Chen and Siau (2012) investigates business intelligence (BI) from the perspective of an organization's ability to detect and respond, in a timely manner, to market opportunities and threats – what the authors call organizational agility (OA – defined in the next paragraph), which also greatly resembles the definition of OR used in this research. In particular, the authors consider BI and information technology (IT) infrastructure flexibility as two major enablers of OA in the business context. The authors chose OA as a dependent variable in their study because of their desire to show the strategic values of independent variables (BI and IT infrastructure flexibility) and the strategic importance of OA (Chen & Siau, 2012, pg. 3).

Chen and Siau (2012, pg. 2) define organizational agility as follows: 'OA is the ability to sense and respond to market opportunities and threats.' The authors then further qualify this definition by stating that 'there are two source components that can help improve organizational agility: (1) the component that can help sense and detect market opportunities and threats in a timely manner, and (2) the component that can help act on or respond to market opportunities and threats in a timely manner.' The definition of OA presented above resembles the definition of OR quoted in this thesis (Robb 2000; Hamel & Valinkagas 2003; Starr et al. 2003; iJet International Inc. 2008). For the purpose of their research, Chen and Siau (2012, pg. 4) refer to the qualities required for IT infrastructure flexibility as connectivity, compatibility and modularity.

Given the similarities between the definitions of OA and OR, the following statement, quoted from the work of other researchers, is of special value to this thesis, as it possibly links OA to OP through the common variable of performance: 'There is an established positive link between organizational agility and firm performance in the IS literature (Benaroch 2002; Sambamurthy et al. 2003; Fichman 2004; Benaroch et al. 2006)'. The importance lies in the fact that OA and OR appear to be defined in very similar ways, so one could possibly assume that, by linking OA to performance, one can also link OR to performance.

Of special importance to this research, due to the similarities between OA and OR, is the hypothesis development and the postulation that 'BI can enhance an organization's agility' (Chen & Siau, 2012, pg. 6). Organizational agility can be facilitated by BI through 'detecting customer event patterns, identifying operational opportunities and bottlenecks, and revealing changing in partners' assets and competencies to managers so they sense, act, or make timely decisions' (Chen & Siau, 2012, pg. 6).

In a study by Chen and Siau (2012, pg. 13) that focused on empirical tests of the contributions of BI use to OA, the authors find support for their hypothesis, as they state '[t]his finding provides the first empirical support that business intelligence has strategic values. Business intelligence should be treated as a critical component of an organization because of its contribution to organizational agility.'

The limitations of this study, as identified by the authors (Chen & Siau, 2012, pg. 14), include its cross-sectional nature, which measured the effects of BI on OA at one point in time. The authors suggest that another study be performed that investigates the impact on the dependent variable over time.

Important to this work is the comment made by Chen and Siau (2012, pg.1) that 'empirical studies on BI are still scarce in academic research.' It is the intention of this thesis, therefore, to contribute to the academic body of empirical work.

Luo et al. (2012, pg. 186) state that the capabilities within the organizational context can be defined as three value disciplines: operational excellence (competitively pricing products or services and delivering them without difficulty or inconvenience), customer intimacy (cultivating relationships with customers and satisfying their needs) and product leadership (offering leading-edge products and services).

Because products and services can quickly become obsolete or be replicated in today's business world, Luo et al. (2012, pg. 186) state that 'firms need to be able to respond consistently and quickly to changing markets. Identifying, acquiring and accumulating critical organizational capability in line with the three value disciplines are critical to firms' "strategic renewal."' The similarities between the description offered by Luo et al. of organizational capabilities based on three value disciplines and OR make their work valuable, as it may provide additional insight into the understanding of OR.

Investigating the relationship between IT assets and organizational capability, Luo et al. (2012, pg. 188) acknowledge the role of analytics/business intelligence, as they state that '[e]xtensive deployment of analytic (business intelligence) systems, for instance, allows firms to access and analyze market and customer data'. That action, presumably, creates knowledge that can be used in products, services and marketing, among other things, in order to achieve competitive advantage. Also worth mentioning is the fact that Luo et al. refer to analytics as business intelligence and vice versa. Such interchangeability of terms between analytics and BI is very common in a business environment where, perhaps due to limited understanding of both, people tend to apply the terms interchangeably.

Luo et al. (2012, pg. 188) mention the following specific areas where analytics can positively affect operational efficiency: customer segmentation, cause-and-effect marketing analysis, market sensing, pricing scenarios and promoting actions.

The work of Popovic et al. (2012) addresses the issue of BI success and is similar to the works of Chen and Siau (2012), Isik et al. (2013), Shallo and Galliers (2013) and Popovic et al. (2012, pg. 730) in that they define success in non-material terms, but they do acknowledge that the most successful organizations mainly focus on capturing the value of information/knowledge throughout the various stages involved in the processing of information and its use. The business intelligence system (BIS) success model proposed by Popovic et al. (2012, pg. 730) consists of the BIS maturity component directly impacting two other components: the information quality (IQ) component and the information access quality (AQ) component. The outputs from IQ and QA, affected by the analytical decision-making culture, flow into the 'measure of BI success' component: the use of information in business processes.

The research design and methodology used by Popovic et al. (2012, pg. 733) is of special interest for this research, as there are many similarities, which are listed below between the work of Popovic et al. and this work; these similarities validate, to a certain extent, the approach chosen for this research.

These similarities include the following:

- To ensure the content validity of the research instrument, Popovic et al. base their questionnaire on a theoretical foundation. The questions used in this research are derived from McKenzie and van Winkelen's 2004 book;

- To ensure face validity, both studies conducted pre-testing of their respective instruments;
- In the case of both studies, the participants were given introductory letters that explained the aims and procedures of the study;
- Measurement items (dependent variables) were developed based on the literature reviews;
- Organizations selected for the analysis were chosen from bodies that aggregate information about organizations. In the case of Popovic et al., the source was the Agency for Public Legal Records and Related Services (in Slovenia). In the case of this research, the provider of organizational information is the Hoover Company, the sister company of the Dunn & Bradstreet Company, which is well-respected for the business-related services it provides;
- In both works, the recipients of the questionnaire were the senior managers;
- Both studies followed the approach of Prajogo and McDermott (2005) by discounting the returned responses as undeliverable;
- In both studies, the pre-test questionnaires were added to the body of general replies due to a low overall return rate; and
- In both studies, Cronbach alpha was used to confirm the construct's validity.

The study conducted by Isik et al. (2013, pg. 13) examined the role of the decision environment in how well BI capabilities are leveraged to achieve BI success.

The findings of Isik et al. (2013, pg. 21) are surprising in one aspect, namely data quality. While, as predicted by the authors and the research model, quality of user access, flexibility and integration with other systems positively affect BI success, data quality is negatively related to BI success, regardless of the decision environment. Iskik et al. (2013, pg. 21) offer possible explanations for this; one of the explanations offered is that data quality in today's BI initiatives is 'good enough,' meaning that additional improvements in data quality may come at the expense of other BI capabilities (Isik et al.,2013,pg. 21).

Kowalczyk et al. (2013, pg. 1) conducted a literature review to determine if the reported business intelligence and analytics success cases with regards to organizational performance have a similar effect when considering the impact of BI&A from the decision process perspective. As stated by the authors, the context for the research was as follows: [t]he realizable benefits from such decision supporting technologies depend on their effects on organizational decision processes' (Kowlaczyk et al., 2013, pg. 1).

For the purpose of this thesis, an attempt to carry the argument further – to seek a correlation between OR and organizational decision processes – is not feasible, as the correlation cannot be convincingly stated; Kowalczyk et al. (2013, pg. 2) cite the work of Shollo and Kautz (2010) in this regard, stating '… very few studies address decision processes and it often remains unclear how BI is used in decision processes and what effect it has on decision processes.' This inability to correlate organizational decision processes with OR excludes the possibility of seeking to determine how the work of Kowalczyk et al. (2013) can help understand how BI&A impact OR through organizational decision processes, which are, as shown in Table 3.4.2.1, a key BI direction.

As the result of their structured literature review, which investigated the effects of BI&A on phases, characteristics and the outcomes of decision processes, Kowalczyk et al. (2013, pg. 10) found that, at the high-level view of decision processes, there were five studies that provided evidence in support of the general perception that a DSS has a positive impact on decision processes. Thirty-two studies that investigated effects more specifically related to the decision processes in more detail found less clear support, prompting the researchers to call for the research to be expanded.

In terms of limitations and issues related to what BI systems can do for organizations, Shollo and Galliers (2012, pg. 10) found that poor quality of data is one of the major inhibitors of the use of BI, which is somewhat contrary to the findings of Isik et al. (2013) presented earlier. In addition, the ways in which BI systems are implemented and how they are used makes a significant difference to their usefulness. This is because it is important to be able to model the BI system in a way that reflects the 'real-life' system closely, as it is crucial to be able to ask and find the answers to the right questions. Finally, Shollo and Galliers (2013, pg. 10) note that frequent changes in the strategic focus of various departments and/or entire organizations can lead to a fragmented view and situations in which the findings obtained from the BI system are not applied due to shifts in priorities.

However, the fact that the case study involved a single company located in one region of Europe is a significant limitation of the Shollo and Galliers study (2013). This is especially problematic from the perspective that emphasizes human aspects in the process of knowing, as different parts of the world may respond to and use BI systems very differently, given their culture and world view.

While Shollo and Galliers (2013) do not imply a direct effect of BI on knowing in organizations, it is crucial in facilitating knowing in organizations through its role in discussions, negotiations and reflections.

Corte-Real at al. (2014, pg. 175) report that successful initiatives within the healthcare, airlines, financial services and telecommunications industries have been studied qualitatively; however, the authors also state that the implementation of BI within an organization does not necessarily lead to improved performance, as a number of companies incurred sizable losses as a result of BI&A initiatives (Corte-Real at al., 2014, pg. 175). Corte-Real at al. (2014, pg. 175) summarize the findings of their literature review, stating that BI&A provided some benefits in the form of increased sales and customer satisfaction, support for strategic decisions, the mitigation of contagious bank failures and the discovery of fraud patterns. Corte-Real at al. (2014, pg. 176) summarize their research by stating '[d]espite a long-standing research tradition investigating the role of IS in decision-making, there is little understanding of how BI&A systems may effectively be used and create positive impacts on the organization.' This quotation further illustrates the need for the research such as that undertaken in this thesis.

3.4.4 DM as an Analytical Tool for Measuring the Impact of KM on OR

This section reviews the literature that focuses on discussing the application of DM, either with respect to KM, OR, or both. The emphasis of this review is on the applicability of such use to the aim of this research, specifically determining if DM is an appropriate analytical tool for measuring the impact of KM on OR. Many of the concepts from the authors presented in the previous section (3.4.3), which focused on theoretical perspectives, are re-visited here, presented in an applied form. In other words, when applying analytical terminology, this section considers the texts that addressed practical aspects of using DM to measure the impact of independent variable (KM) on the dependent variable (OR).

3.4.4.1 Impact of DM on KM

The review of the practical applications of DM with respect to KM leads to the following findings:

The problems related to the relatively low applicability for business organizations of problems solved by DM have been listed by Hopkins and Schadler (2015, pg. 10) as one of the key three issues encountered when transforming data into actions as they state: 'poor linkage between insights discovery and business action and scarce learnings from actions taken.' Hopkins and Schadler's view has been shared by many writers recently, including Cao and Zhang (2006), Brusilovski and Brusilovski (2008), Adejuwon and Mosavi (2010), Wu et al. (2010), Li et al. (2012), Shollo and Galliers (2013), Corte-Real et al. (2014), Hopken (2014) and Rao (2015).

Cao and Zheng (2006, p. 49) present a practical DM methodology that allows for the generation of actionable knowledge in a constraint-based environment, commonly referred to as domain-driven data mining (DDDM), a concept introduced in Section 2.5.6. Within the DDDM methodology, Cao and Zhang (2006, pg. 53) present the domain-driven in-depth pattern discovery (DDID-PD) framework. Cao and Zhang (2006, pg. 50) also emphasize the importance of the involvement of domain experts and the knowledge they possess in the development of effective data mining techniques for business organizations. This 'domain expert emphasis' is reflected in the DDID-PD model.

Cao and Zhang have successfully applied the DDID-PD framework in mining actionable correlations in the Australian Stock Exchange, thereby showing the potential of DDID-PD for improving the actionability of extracted knowledge for practical use by businesses (Cao & Zhang, 2006, pg. 49).

According to Moayer and Gardner (2012, pg. 70), SKM, introduced in Section 3.4.3.1, can overcome problems typical of project-based industries (like mining) that primarily rely on matrices superimposed on functional structures in order to align capacity, personnel expertise and business requirements. Strategic knowledge management overcomes this weakness by using clearly articulated organizing principles to drive KM and OL activities.

In one of the very few works that quantifies the impact of BI on organizations, Shollo and Galliers (2013, pg. 2) quote the work of Brynjolfson et al. (2011), which 'provides evidence that the adoption of BI technologies leads to a productivity increase of between 5 and 6%.' However, Shollo and Galliers (2013, pg. 2) note that very little research on the role the BI systems play in organizational decision-making had been performed by the time of their paper. A similar finding was stated by Kowalczyk et al. (2013, pg. 2), who cited the work of Shollo and Kautz (2010) in this regard: 'very few studies address decision processes and it often remains unclear how BI is used in decision processes and what effect it has on decision processes.'

Shollo and Galliers (2013, pg. 5) conducted an interpretative study that sought to determine how BI systems mediate knowing in organizations as a pre-cursor to managerial decision-making. Two main concepts emerged from their work, namely data selection and articulation. Data selection provides a fresh look at a situation many times in multiple dimensions, while articulation represents an opportunity to articulate a hypothesis based on results obtained from the BI. The human aspect involves the dialogue intended to make sense of the results obtained from BI that takes place among members.

While Shollo and Galliers (2012, pg. 9) appear not to discuss predictive or prescriptive uses of BI, perhaps due to the less than widespread use of such analytical approaches in the marketplace at the time the research was conducted, they do refer to the less emphasized drill-down, roll-up capability of the BI systems, stating that 'the capability of BI systems to enable people to drill-down and roll-up data, enables them to track the data at each step, thereby facilitating discussion about the assumptions underpinning the analysis, which leads to better understanding of other perspectives.' From the author of this research's professional experience, the drill down/aggregation capability, from a DM practitioner point of view, carries little value when compared to the possible outcomes obtainable from the use of predictive analytics, as such capability represents the lowest (descriptive) level of DM.

The work of Fuchs et al. (2014) is of special interest for this thesis, as it describes the practical application of business intelligence and analytics (BI&A) in the creation and application of knowledge for the purpose of improving business at tourist destinations, along with support for the superiority of the application of DM over traditional statistics for solving certain business problems. Its measurement of an intangible (tourist satisfaction) greatly adds to the importance of the paper. The work of Fuchs et al. builds on the prior published work of all authors, which serves as its theoretical foundation (Hopken et al., 2011), and it presents the practical 'knowledge destination framework' and the 'knowledge destination architecture'.

Fuchs et al. (2014, pg. 199) refer to the concept of learning tourist destination introduced by Schianetz et al. (2007); the concept resembles, in many ways, the focus of this study, namely organizational resilience. Schianetz et al. suggest that 'the learning focus should be the understanding of how tourism destination functions, how market possibilities can be enhanced, the requirements for application of changing environments, how to promote collective awareness of economic, social and environmental risks and impacts, and how risks can be minimized' (2007, pg. 1486). Similarly, the qualities mentioned above tend to be those used in defining OR (Mallak (1998); Robb (2000); Hamel & Valinkangas (2003); McCann et al. (2009)).

Based on a literature review and past work, Fuchs et al. argue that 'knowledge creation and acquisition processes at tourist destinations can be significantly enhanced by applying methods of Business Intelligence (BI)' (2014, pg. 199), where the methods of BI, according to Fuchs et al., consist of source data identification, ETL and DM processes (2014, pg. 199).

The knowledge destination framework presented by Hopken et al. (2014) consists of four quadrants, which split the framework into four categories: the customer oriented and supplier oriented (vertical split) categories and, split horizontally, the knowledge application and knowledge generation categories (Fuchs et al., 2014, pg. 200). The knowledge activities focus on the extraction of information from customer-based and supplier-based sources as well as on the generation of knowledge that could be applied to improve business and customer satisfaction. In essence, DM is one of the key components used to generate knowledge for the sake of the improvement of business and customer satisfaction, which should further improve business opportunities.

The creation of knowledge on the customer-side of the model involves the use of surveys, social media and/or web-based comments, as well as GPS-provided tourist locations. The creation of knowledge on the supplier-side includes the use of DM for customer profiling and products, among other activities. The application layer on the customer-side consists of various customer-based services, which include recommendations and location-based offerings, whereas the supplier-based application involves suppliers' access to the knowledge bases and data visualization (Fuchs et al., 2014, pg. 200).

As a specific example of the application of DM and the creation of knowledge in the tourism industry, Fuchs et al. (2014, pg. 200) state that DM methods of supervised learning and estimation can be used to account for tourist bookings and cancellations in order to predict tourist demand. In addition, unsupervised learning-based DM algorithms enable the segmentation of customers for a better understanding of the composition of various customer groups (Fuchs et al., 2014, pg. 200). The use of both supervised learning (based on the known answers to the questionnaire in the OR section) and unsupervised learning are very informative for this research.

Based on their studies Fuchs et al. (2014, pg. 204) state that the greatest knowledge creation occurs in the 'post-trip' phase where tourists provide, via surveys, feedback related to their experiences. The tourists' replies are later compiled and mined in order to learn about their future needs and preferences.

New knowledge is also generated by applying, among others, Naïve Bayes and nearest neighbor algorithms that allow text mining from social media platforms such as TripAdvisor and Booking.com, where positive, negative and neutral statements about particular tourist destinations or attractions are retrieved from social media. These experiences are later analyzed, aggregated and presented in a dashboard form (Fuchs et al., 2014, pg. 205).

In their work, Leung and Joseph (2014, pg. 710) present, as they call it, a sports data mining approach that helps to provides interesting knowledge about sports games and predict their outcomes, with a focus on college football. The uniqueness of the approach chosen by Leung and Joseph lies in the fact that their prediction model does not fall into one of the two quantitative prediction model categories that are currently widely used, which are, respectively, simulation-based (which makes predictions through the use of simulation engines), or statistics-based (which relies on the statistics of the teams competing) (2014, pg. 712). Instead, Leung and Joseph researched the approach of predicting the outcome of a football game by analyzing the teams that are the most similar to each of the competing teams, finding the results of the games between those teams and, finally, using the outcomes of those games to predict the outcome of the game between the original teams (2014, pg. 716). While Leung and Joseph do not appear to use any specialized data mining algorithms and instead carry out various computations in a traditional manner, their approach to pattern-finding appears to be in line with the approach adopted by the commercial data mining algorithms, which search for patterns that are later used for prediction.

The work of Natek and Zwilling (2014) illustrates the successful use of data mining solutions in an attempt to answer the following research question: ‘Are there any specific student characteristics, which can be associated with the student success rate?’ (Natek & Zwilling, 2014, pg. 6400). In carrying out their research, the writers were also interested in determining if DM has the potential to become a serious part of the knowledge management system of higher education institutions, in terms of assisting professors and researchers with decision-making (Natek & Zwilling, 2014, pg. 6406).

In the process of selecting DM tools to carry out their research, Natek and Zwilling (2014, pg. 6402) investigated various tools offered by the Microsoft Corporation. The authors state that Microsoft, at the time of writing, offered three levels of data mining solutions:

- The basic level of DM, which includes tools such as the Excel spreadsheet program;

- The intermediate level of DM, which includes DM extension tools for Excel (which are free of charge add-ons provided by Microsoft); and
- The advanced level of DM, which includes DM tools included in Microsoft's flagship database product, SQL Server.

As noted by Natek and Zwilling (2014, pg. 6402), all of the above DM levels use algorithms from Microsoft SQL Server but can be distinguished by their user interfaces and the different techniques and parameters used to manage the data mining process. For their research, Natek and Zwilling (2014, pg. 6402) chose the basic level; however, for the purpose of this research, the advanced level was chosen as better matching the interest and computational abilities of the author.

Perhaps the most appealing aspect of the research conducted by Natek and Zwilling was their conclusion regarding the application of DM to a very small set of data (2014, pg. 6402), as a very similar challenge (using a small data set in DM) was also encountered in the research conducted for this thesis. While Natek and Zwilling state that it is a well-known fact in the industry that data mining algorithms work best with large data sets, they find that small student data sets did not limit the use of the mining tools when performing specific data mining analysis (2014, pg. 6404). Using a data mining tool called WEKA in addition to Excel, Natek and Zwilling also found that several decision tree algorithms are very practical when working on small data sets (2014, pg. 6406).

The authors also acknowledge that not all analyses are possible for every data set; they acknowledge that forecast, scenario analysis and shopping basket algorithms may not be applicable due to limitations in data or content (Natek & Zwilling, 2014, pg. 6402).

The work of Natek and Zwilling (2014) resembles this research work in numerous dimensions, which are identified below:

First, the research question posed in their paper highly resembles the research question addressed by this research, as both seek an association between independent and dependent variables;

Second, the steps taken to arrive at the results were similar in both cases: creating a model data set and choosing data mining technology and techniques (Natek & Zwilling, 2014, pg. 6402);

Third, both works used DM tools provided by the Microsoft Corporation; and

Finally, both works were challenged by relatively small data sets that served as the basis for input data.

Lamont (2015a, pg. 9) describes an analytical product offered by a company called ClickTale (the product is also named ClickTale) that, in addition to offering an analytical engine, uses psychologists as a part of the service provided to conduct further analysis, based on customers' actions on a particular web site. Lamont (2015a, pg. 9) quotes ClickTale's statement regarding their product: 'Using information from analytics, we try to understand a visitor's motivations and decision-making process'.

Finally, Rao (2015, pg. 3) shares several takeaways from the KM Singapore 2015 conference that are important for the purpose of this thesis. The focus of his article is on the factors that, once properly applied, can contribute to an organizational advantage that is built on knowledge. The key factors that align well with this work include the following:

- Because of today's fast-paced business environment, KM needs to be able to keep pace with business changes and be able to demonstrate quick wins;
- Tying knowledge to learning by promoting knowledge learning, instead of a knowledge-sharing culture;
- Ensuring knowledge succession so that it can be sustainably maintained; and
- Finally, and most importantly for this work, bridges between KM and DM must be built, as DM can provide useful insights if the right questions are asked; this is where KM can become very valuable.

3.4.4.2 Impact of DM on OR

This section reviews the literature related to the impact of DM on the dependent variable (OR) in the real-world context.

As specific examples of the application of DM, Ngai et al. (2009, pg. 2595) cite Carrier and Povel (2003), stating that DM, by employing segmentation algorithms that group customers based on different characteristics and needs, can improve marketing campaigns by increasing response rates. It can also predict how likely a customer is to leave to a competitor.

Based on their research, Ngai et al. (2009, pg. 2599) identify some areas where data mining can play a large role in making businesses successful. These suggestions include the following:

- Customer complaints management – DM could be used to seek patterns in data relating to customer complaints;
- Root problem analysis (in customer relations) – using DM to analyze associations between complaints from different customers; and
- Target customer analysis – using neural networks and decision trees algorithms (in addition to the classification and segmentation algorithms that are currently commonly used) to identify profitable customer segments through analysis of customer's underlying characteristics.

The work of Morales and Wang (2010, pg. 554) focuses on the area known as revenue management (RM), which, according to the writers, 'enhances the revenues of a company by means of demand-management decisions.' As such, it can be anticipated that proper RM, through methods such as dynamic pricing and capacity allocation, can have a positive impact on a company's bottom line, positively affecting OR.

The work of Morales and Wang (2010) examines employing DM along a the real-world dataset collected over a period of three years in order to model behavior of people staying at hotels and cancellations. The most important contribution from the work of Morales and Wang, according to the researchers (2010, pg. 555), is the fact that their work addresses the modeling of customer behavior at various stages of the booking horizon (from the time of booking that occurs well in advance to the time of t=0, the time of service). The text byMorales and Wang (2010, pg. 556) focuses on what the authors refer to as 'complete cancellation curve.' That is, they seek out models that are capable of modeling cancellation for any (practical) time period t between the booking and the time of service.

With respect to KM, the work of Morales and Wang (2010, pg. 557) focuses primarily on its knowledge creation aspect. As stated by the authors (2010, pg. 557), knowledge about the dynamics of customer cancellations and their dependence on the time-to-service timeframe will help in building more accurate forecasting models as well as in understanding the drivers of cancellations.

Chen and Siau (2012, pg. 5) use the awareness-motivation-capability (AMC) framework introduced by Chen in 1996, suggesting three organizational behavioral drivers: awareness (manifested, among other ways, in action visibility and firm size), motivation (territorial interests in different markets) and capability (execution difficulty and information processing). With respect to the AMC network, Chen and Siau (2012, pg. 5) state that '[b]usiness intelligence can help raise the awareness of opportunities and threats in marketplaces, then motivation of responding follow'.

On the practical side, but perhaps slightly counter-intuitively, when evaluating suppliers using DM, Aghdaie et al. (2014, pg. 774) found that, when evaluating the performance of the suppliers of a large automotive manufacturer in Iran, not the purchase costs but the flexibility (in production, production volume and labor force) were given the largest weight, with purchase costs being second and delivery performance third.

Based on the literature review performed Chae et al. (2014, pg. 121) cites the work of Sharma et al. (2010, pg. 193) suggesting that there may be an indirect relationship between analytics and performance, notes that analytics enables manufacturers 'to apply resources to undertake actions to deliver performance gains and competitive advantage'.

While the research of Chae et al. (2014) supports the positive impact of accurate manufacturing data and analytics on organizational performance, the research does not appear to be looking at other factors possibly affecting organizational performance: the most common one being the market conditions. Given the exploratory type of their research it would be interesting to see what type of results would have been obtained using data mining.

3.4.4.3 Impact of KM on OR and Measurement of Such Impact Using DM

This final section examines the literature review with regard to the impact of KM (the independent variable) on OR (the dependent variable) in the context of real-life scenarios measured using DM.

Due to the very limited number of published works, this section combines the discussion of the impact of KM on OR with the measurement of such impact using DM.

In their paper, Choi et al. (2008) use data on KM and organizational performance collected from 131 Korean public firms to assess the synergistic relationships between KM strategies and the impact of such strategies on organizational performance. As stated by Choi et al. (2008, pg. 237) in their paper, 'complementarity indicates a condition of increasing returns in which adopting (doing more) of an activity (e.g. implementation of certain KM strategy) has a higher payoff when simultaneously adopting (doing more) of a complementary activity (e.g. implementation of another KM strategy).'

In terms of the review of literature addressing the impact of KM on OR as measured by the DM, the following work appears to offer some practical insights.

Wu et al. (2010, pg. 400) state that only a few studies exist that attempt to use DM to measure the contribution of KM to organizational performance and that the contributions of KM, which are mostly qualitative in nature, are hard to measure. While the study did not find specific hidden patterns between KM and ROA (the measure of operational performance used in the research) the outcome of the Bayesian network classifier (BNC), the directed acyclic graph, showed the KM purpose and tacit-oriented degree as the most influential attributes (Wu et al., 2010, pg. 399). While the work of Wu et al. (2010) provides limited insights into the impact of KM on organizational performance, it does provide a new and practical method, using BNC and rough set theory (RST), for the discovery of hidden relationships between KM and organizational performance expressed by ROA. However, the limited number of factors considered in the study, the study's sample size (85 surveys) and the simplicity of the performance indicator call for additional research in the area, as stated by the authors (2010, pg.399). Worth noting is the approach taken by Wu et al., which classifies the values of the dependent variable, the ROA, into two classes: the higher and the lower performing group. The SPSS two-step cluster method was then used to classify the ROA attribute into the two classes resulting in 40 organizations being placed in the higher performing group and 45 in the lower. Given the lack of existing literature that addresses the impact of KM on OR and does so via DM analysis, this thesis seeks to contribute to the body of knowledge.

3.4.5 Summary

The literature review in this chapter focused on reviewing texts related to all of the key themes in this research: KM, OR and DM and the use of DM as an analytical tool to assess the impact of KM on OR. The chapter began by presenting developments in the DM field in the context of KM and OR. Thereafter, the literature review progressed by investigating specific relationships between the key elements of this research.

3.5 Organizational Resilience – Proposed Model.

The purpose of this section is to present, based on the multi-field (KM, OR, DM) literature review conducted in Sections 3.2, 3.2 and 3.4, a theoretical model for the resilient organization: an organization that exhibits very high levels of OR, based on the definitions of OR used in this research and presented in Section 2.4. The purpose of this model is to provide a framework for implementing processes and procedures that facilitate an organization becoming resilient; in addition, the model is presented to fill the gap in the literature regarding the OR framework.

The review of the literature conducted in the previous sections provides the background for the selection of elements to be included in the proposed OR model. The choices of the components of the OR model were driven by an understanding of OR (shaped by the literature review) and the practical experience of the author of this work. The foundation of the model is derived from the work of Moayer and Gardner (2012). The work of these authors was selected as the basis for this model because of the authors' comprehensive approach to OR, as was discussed in Section 3.4. (To briefly summarize, the components derived from the work of Moayer and Gardner include the strategic knowledge management [SKM] framework with four perspectives, the shared organizational principles and the use of DM.) Further improvements to the model, added to allow a more complete view of the world, were taken from the work of the following writers: Robb (2000), due to his work on the performance and adaptation systems; Hamel and Valinkangas (2003), due to their emphasis on organizational culture and the need for environmental sensing; McKenzie and van Winkelen (2004), who provided the elements from the competence and monitoring areas; Cao and Zhang (2006), whose work contributed the constraints associated with DM; and Shollo and Galliers (2013), who enhanced the OR model through the addition of the learning process. The model and its components are presented, with a discussion of the contributions made by the authors mentioned above.

The 'human side' of the model is comprised of strategic knowledge management (a term that, based on the discussion of Moayer and Gardner [2012, pg. 67], means employing knowledge management [KM] as a part of the overall business strategy, where the strategy is the art of formulating, implementing and evaluating the business decisions that lead to the achievement of organizational goals and objectives) and the 'technological side' consists of the data mining system.

The four views of strategy (Moayer & Gardner, 2012, pg. 68) that provide the input to strategic knowledge management (SKM) consist of the following:

- An economic perspective of the world: a market-based view with the goal of achieving a preferred position within the industry;
- A political perspective or stakeholder-based view, with the goal of engaging stakeholders in decision-making in order to facilitate the achievement of business goals;
- An internal human, structural and capital asset ability perspective or resource-based view, with the goal of achieving the best utilization of human, financial, physical and organizational resources; and
- A knowledge-based view that focuses on knowledge creation and utilization for the creation of value for the organization, which is somewhat of an extension of the resource-based view that considers

knowledge as a valuable (if not the most valuable) organizational resource.

The four strategic perspectives described above feed the SKM, with a focus on two systems defined by Robb (2000, pg. 27), which were discussed in Section 3.2: the 'day-to-day' or current goals and the 'adaptation system' or future positioning. According to Robb, both are needed for a firm to be resilient (2000, pg. 29). The dual focus of SKM on both the 'present day' as well as the 'future day' is also in line with the work of McKenzie and van Winkelen (2004, pg. 13; pg. 235). McKenzie and van Winkelen state that in the competing competence area, there is a need to pay attention to both knowledge exploitation and knowledge exploration, and, in the monitoring competence area, there is a need to monitor the current performance of the value of knowledge as well as developing knowledge in order to be able to adapt to change. Moreover, the monitoring competence area of McKenzie and van Winkelen (2004, pg. 235) provides a governance mechanism between the current goals and future positioning within the OR model.

The DM 'side' of the model is responsible for extracting knowledge from the data collected by an organization as well as from the external sources brought into data storage, such as a data warehouse. Moayer and Gardner (2012, pg. 69) note that the strategic strength of DM, in their model, is derived primarily by solving unstructured business problems, solutions to which are hard to replicate due to business and environmental differences among different organizations. As discussed by Hamel and Valinkangas (2003, pg. 12), the need for environmental scanning in order to detect threats to an organization is also fulfilled by the DM component through the use of predictive modeling.

Data mining activities and models are also subject to domain and data constraints, as suggested by Cao and Zhang (2006), to reflect the fact that real-life business problems are subject to such constraints. (The constraints presented by Cao and Zhang [2006, pg. 50] were discussed in Section 2.5.6 and 3.4.3.2.)

Besides acting as an integrating force, DM delivers substantial value through the interactive learning process that combines heuristic questioning, organizational learning routines, knowledge assets, sense making and strategy in order to interpret market signals with relation to the current goals and future positioning. Data mining can serve as a balance between a subjective and an objective interpretation of business goals. To further enhance the interactive learning processes and improve business outcomes, the articulations suggested by Shollo and Galliers (2013, pg.7) should to be used. Shollo and Galliers (2013, pg. 7) define articulation as 'the coherent communication process of one's beliefs, opinions and ideas.' The articulations should include i) supplementing data and results with personal knowledge to provide appropriate context (similar to the DDDM concept discussed in prior section and in Section 2.5.6); ii) analyzing how the extracted knowledge affects an organization at various organizational levels; and iii) deciding which DM's results should be pursued.

Interactive learning takes place within the model's 'shared organizational principles' specified in Moayer and Gardner (2012, pg. 70), which are embedded in management routines that align human-technology interactions and organizational structures, norms and values. Added to the shared organizational principles, and also guiding the strategy, learning process and decisions, is the component of organizational culture that supports resilience. This addresses, as pointed out by Hamel and Valinkangas (2003, pg. 3), four challenges that stand in the way when becoming resilient. (The four challenges – cognitive, strategic, political and ideological – were discussed in Section 3.3.2.)

While this model has not been applied in this thesis, given the aim of its research, it is expected that the application of the model in an industrial setting would lead to greatly improved OR; the level of OR could be measured prior to and after the implementation of the model by the methods presented in this research (in Chapter 6).

The proposed OR model for an organization can be summarized as follows.

The market-based view provides an economic context for understanding the marketplace and the positioning of an organization in the market. The stakeholder-based view seeks to engage the stakeholders in the business for the purposes of intelligence gathering and decision-making. The resource-based view ensures that the resources required by the organizational strategy are available. The knowledge-based view seeks to create value for an organization based on the knowledge that it possesses.

The four views provide support for the business strategy of an organization that includes KM as a key component and that employs a strategy that focuses on both current goals and future opportunities (as identified by the knowledge generated by DM, which operates under real-life constraints).

Finally, all of the actions mentioned take place with the guidance of well-defined and shared organizational principles, supported by an organizational culture that facilitates learning and encourages warranted adjustments to the business strategy.

The proposed model of the resilient organization, therefore, has the following appearance:

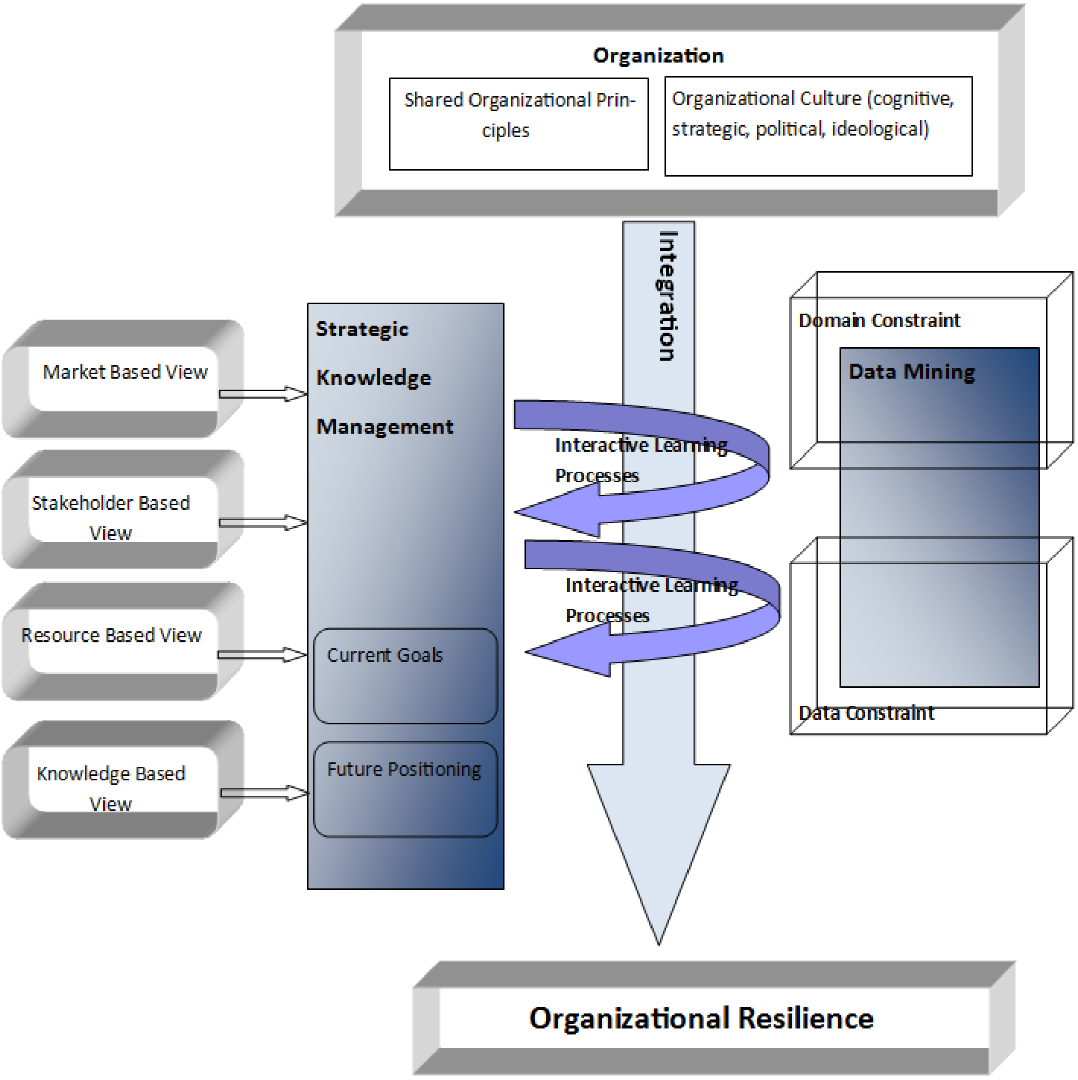

Organization
Shared Organizational Principles
Organizational Culture (cognitive, strategic, political, ideological)
Integration
Market Based View
Stakeholder Based View
Resource Based View
Knowledge Based View
Strategic Knowledge Management
Current Goals
Future Positioning
Interactive Learning Processes
Interactive Learning Processes
Domain Constraint
Data Mining
Data Constraint
Organizational Resilience

Fig. 3.5.1: The proposed OR model

3.6 Summary of the Literature Review

3.6.1 Introduction

This chapter summarizes the literature review conducted in this research and focuses on addressing research questions #1 and #2. In addition to providing the answers to the research questions identified, this chapter also identifies any gaps found in the literature that is applicable to this research. One additional point needs to be made regarding direct references to OR in the context of KM, which is that such references may not actually be available. The discussion about OR in the context of KM is often carried out through the concepts, identified in the literature review, that are the 'next best OR substitute' namely the concepts of organizational performance (OP) and competitive advantage (CA). The principle discussion about the suitability of OP and CA to define OR takes place in Section 3.4.3.1, and the issue is therefore only highlighted in this chapter. The gaps in the literature and manner in which this research has sought to fill those gaps close the discussion in this chapter, and the summary chapter follows. The conclusions of the entire research project are presented in Chapter 7.

The layout of this section is shown in Fig. 3.6.1, below:

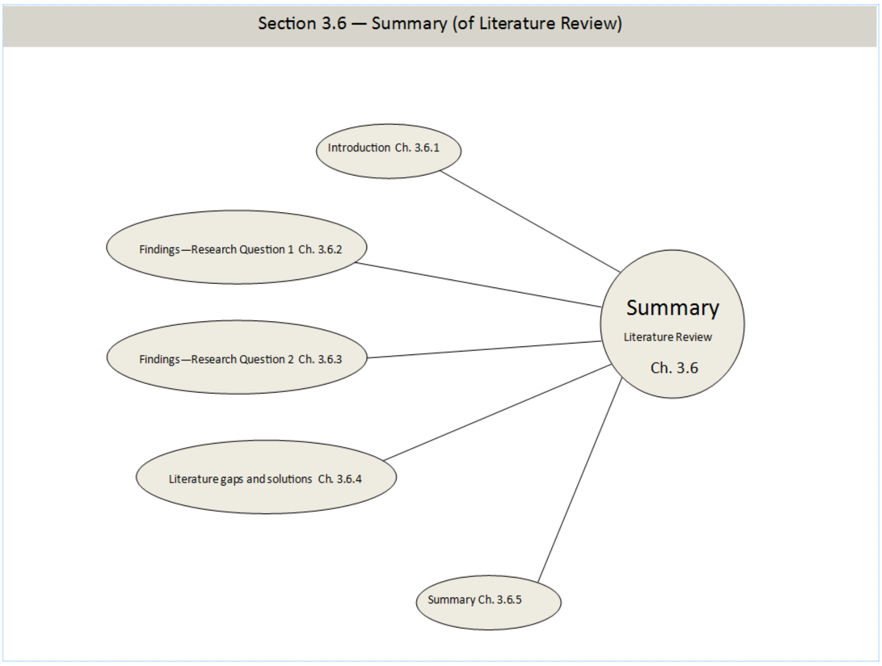

Section 3.6 — Summary (of Literature Review)
Introduction Ch. 3.6.1
Findings—Research Question 1 Ch. 3.6.2
Findings—Research Question 2 Ch. 3.6.3
Literature gaps and solutions Ch. 3.6.4
Summary Ch. 3.6.5
Summary
Literature Review
Ch. 3.6

Fig. 3.6.1.1: Structure of the summary of literature review section

3.6.2 Findings in Regards to Research Question #1

This section of the chapter addresses research question (RQ) # 1 (presented in Table 3.6.2.1) and, while it builds on the entire literature review, the contents of this section are most closely related to Section 3.4.3, which examined the theory-based aspects of the evaluation of KM and OR and the role of DM in such evaluations.

When reviewing the literature from the perspective of a process-based view of KM, it appears, perhaps not surprisingly, that the primary knowledge management processes discussed in works examining the impact of DM (also referred in this section to as BI&A) on KM are knowledge creation and utilization (Cao & Zhan (2006), Ngai et al. (2009), Corte-Real (2014), Leung & Joseph (2014), Chemchen & Drias (2015), Hopkins & Schadler (2015)), and, to a lesser extent, knowledge storage and retrieval (Lee, 2008). While considering the role of KM in answering RQ #1, a recent publication (Gehl, 2015) concerning knowledge sharing in KM processes can perhaps justify the absence of this process within the most documented KM processes in the context of DM. The sharing of knowledge generated by DM, a key knowledge process, may actually not take place at all according to Gehl (2015), who promotes the view of knowledge as a valuable corporate commodity and/or resource (supporting a resource-based view), which is not as easily shareable as any other resource possessed by a corporation.

Research Question #1: What prior research exists in relation to the application of DM with respect to KM and OR and the impact of KM on OR, and what are the known relationships between KM and OR?	Objective: To determine the feasibility of using DM when evaluating KM, OR and/or the impact of KM on OR. Also, to determine the applications of DM techniques that have been developed in support of KM and OR as well as to identify the	Methodological Approach: First round of the literature review focuses on the fields of KM and OR. Then, a second round of the literature review focuses on examining the impact of KM on OR through the lens of DM. (Overall, the literature reviews

	areas of convergence between DM, KM and OR.	employ a circular approach.) Mapping of theoretical and practice-based research in DM, KM and OR from the literature review.

Table 3.6.2.1: Research question #1

Knowledge creation and utilization are the two processes that appear most often in the DM literature when reviewing the publications concerning DM's relation to KM. A number of writers (Cao & Zhan (2006), Ngai et al. (2009), Moayer & Gardner (2012), and Fuchs et al. (2014)) view DM as an explicit tool for the generation of practical knowledge that, when utilized, ultimately leads to organizational benefit.

One group of writers appears to be the strongest advocates for DM-generated knowledge, linking the knowledge generated by DM to competitive advantage. The competitive advantage obtained through the use of DM and created by DM-generated knowledge is discussed by the following authors: McKenzie and van Winkelen (2004), Green (2006), Brusilovski and Brusilovski (2008), Adejuwon and Mosavi (2010), Shih et al. (2010) and Brown et al. (2011). The strategic view of knowledge is also shared by the KM writers Gupta and McDaniel (2003).

Not all writers are ready, however, to associate the use of knowledge created by DM with acquiring competitive advantage. Some writers (Lee (2008), Ngai et al. (2009), Kowalczyk (2013), Chae et al. (2014) and Hopken (2014)) stop short of claiming that such knowledge improves operational performance and leads to competitive advantage. These writers claim that the impact of DM and knowledge is somewhat limited and only state that DM and the knowledge generated by it improve operational performance. The work of Murray (2002) in the area of KM that considers knowledge an enabler of actions leading to the improvement of organizational performance also supports this view. The McKenzie and van Winkelen (2004) model for leveraging knowledge resources mentions both the improvement of operational performance and competitive advantage as goals.

Somewhat related to the views of the writers expressed above, yet sufficiently different to warrant their own classification, are the views of several writers, including Lee (2008), Adejuwon and Mosavi (2010), Li et al. (2012), Popovic et al. (2012), Kowalczyk et al. (2013), Leung and Joseph (2014), Natek and Zwilling (2014), Chemchen and Drias (2015) and Lamont (2015a). These authors view DM and the knowledge created by the mining processes as an enhancer of quality analytical decision making, thus affecting the performance of an organization.

Other KM writers who promote the concept of the utilization of KM, which is the most documented aspect of DM outcomes, for the purposes of improving organizational performance, gaining competitive advantage and improving efficiency include Carlucci and Schiuma (2006), Vorakulpipat and Rezgui (2008), Ibrahim and Reid (2009), West and Noel (2009) and Crook (2011).

Another group of writers offers practical insights and real-life applications of DM in knowledge creation and utilization, from the practical use of the BI tool to the practical discussion of deriving value from analyzing tourist hotel bookings and hotel capacity. Such practical information is shared by the following writers: Morales and Wang (2010), Brown et al. (2011), Tsai (2013), Chae et al. (2014), Corte-Real (2014), Hopken (2014), Leung and Joseph (2014), Natek and Zwilling (2014), Hopkins and Schadler (2015) and Lamot (2015-A).

Yet, despite the very positive impact of DM mentioned above, one of the key issues repeatedly identified in the literature is the human aspect. A number of authors recognize the need for context when working with DM algorithms and models. While the DM algorithms are flawless in carrying out numerical computations, they are unable to present the findings in any given context. The following authors discuss the need for humans in the interpretation of DM results and/or knowledge creation in a business context: Adejuwon and Mosavi (2000), Cao and Zhan (2006) and Shollo and Galliers (2013).

As pointed out in the introduction to this chapter, the concept of OR, as defined for the purpose of this research in the context of measuring an impact of KM on

OR, has not received a great deal of attention in the literature. (This issue was discussed in Sections 3.2 and 3.4.) A two-step process has been used to evaluate the impact of KM on OR. In the first step, there was a need to establish the argument that the number of shared features known from academic research on the impact of KM on organizational performance, namely efficiency, effectiveness and competitive advantage, match the features of OR as defined for this research project. Such a comparison is supported by the OR work of the following authors, which is discussed in Chapter 3: Horne (1997), Mallak (1998), Robb (2000), Hamel and Valinkangas (2003), Starr (2003) and iJet (2008).

In the second step, KM's impact on OP/CA (and therefore, according to the paragraph above, on OR) was successfully established. The impact of KM on OP/CA in this manner is supported by the work of Venzin et al. (1998), Armistead (1999), Yli-Renko et al. (2001), Gupta and McDaniel (2002), Hussain et al. (2004), McKenzie and van Winkelen (2004), Anonymous (2006), Frappaolo (2006), Fink and Ploder (2007), Ibrahim and Reid (2009), Vatafu (2011), Crook et al. (2011) and Chou (2011). These works are discussed throughout Chapter 3.

While in the paragraph above the works cited focused more on KM in the KM-OR relationship, the writers who focus more on the OR side of this relationship also identify the positive impact that KM has on OR. The works of Lee (2008), Choi et al. (2008), Moayer and Gardner (2012) and Gehl (2015) support the notion that KM has a positive impact on OP/CA (and therefore OR, given the similarities of the definitions already discussed in this section). The discussion of these works was presented in Section 3.4.3.2.

The impact of DM on KM has been addressed in different ways by writers. Cao and Zhang (2006) propose a new framework for the generation of practical knowledge for business needs (the use of which leads to improved organizational performance). Similarly to Cao and Zhang, Wang and Wang (2008) propose their own framework for the integration of the knowledge discovery offered by DM with KM. Shollo and Galliers (2013) take a more knowledge-based perspective, stating that BI serves as a balancer of objectivity and subjectivity given that subjective insights and tacit knowledge are articulated using methods such as the backing up of BI results, making it more acceptable and appreciated.

Additionally, some writers attempt to use novel concepts for improving the generation of knowledge by DM, such as the concept of extenics used by Li et al. (2012), described in Section 3.4.3.2.

Finally, more emphasis has recently been placed on the fact that there is simply too much data and too little actionable knowledge generated by DM from the data (Ngai (2009), Hopkins & Schadler (2015)). The solution proposed by Hopkins and Schadler (2015) is the use of machine learning for sense-making. This is a very different view from that of the writers who suggest that human experts should be employed in order to make sense of the results of DM. The approach of using experts for making sense of the outcome of DM process has been mentioned by many writers: Cao and Zhang (2006), Brusilovski and Brusilovski (2008), Adejuwon and Mosavi (2010), Wu et al. (2010), Li et al. (2012), Shollo and Galliers (2013), Corte-Real et al. (2014), Hopken (2014) and Rao (2015).

The impact of DM on OR (again, indirectly through the similar concepts of OP and CA) is perhaps one of the better documented aspects of this research.

A number of writers see DM, and the useful information it generates, as a source of competitive advantage: Brusilovski and Brusilovski (2008), Adejuwon and Mosavi (2010), Brown et al. (2011), Chen and Siau (2012) and Luo et al. (2012).

Brusilovski and Brusilovski (2008) present a very compelling view of the impact of DM on OR. They state that DM is not the aspect responsible for acquiring CA (due to the fact that such tools are available to everyone); rather, gaining CA relies on how the results of DM are interpreted, the software solution used and how creative people apply the knowledge gained from the DM.

Some writers (Luo et al. (2012), iJet (2008)), similarly to the KM-based writers whose work was discussed earlier (Robb (2000), Starr et al. (2003) and Hamel & Valinkangas (2003)), stress the importance of DM in detecting changes in the environment in order to be able to properly respond to and handle them.

Because of the similarities between the two pieces of research, particularly in terms of the dependent variables used, the work of Chen and Siau (2012) is important to this thesis. Chen and Siau (2012) use the concept of organizational agility (OA, being the ability to sense and respond to environmental change) as the dependent variable in their research, which investigates how DM and IT infrastructure affect OA. In the case of this research, OR is the dependent variable and the KM processes (which, in most models, are grouped into McKenzie and van Winkelen's competence areas) are the independent variables that affect OR. At the time they wrote, Chen and Siau stated that there were very few empirical tests of the contributions of DM to OR and that their work was the first such work appearing in publication. This is also similar to this work in that there has also been very little written about attempts to measure the impact of KM on OR through the use of DM.

When examining OR from the perspective of decision-making, Shollo and Kautz (2010), Kowalczyk et al. (2013), Isik et al. (2013) and Corte-Real et al. (2014) correlate better decision-making with OR.

On the less positive side of the application of BI, Corte-Real et al. (2014) point out that BI initiatives often lead to sizable losses being incurred due to expensive and lengthy implementation.

The review of the literature had a profound effect on answering RQ #1 as well as on the design of this research and the principles guiding it. All aspects of the research are discussed in Chapter 4.

3.6.3 Findings in Regards to Research Question #2

The slightly overlapping aspects of RQs #1 and #2 in terms of OR are re-stated below and are addressed in this section. That is, Section 3.6.2 touched on the difficulties of identifying the impact of KM on OR because of the lack of use of the concept of OR as defined for the purpose of this research; yet, based on the prior discussions in this research (Section 3.4 and 3.6.2), it is possible to equate the existing concepts of OP and CA with OR. This chapter now builds on the argument that there are similarities between OP, CA and OR; while the prior section discussed the relationship between these concepts, this section examines the measurement aspect from a practical perspective and focuses on the aspect of measuring the impact of KM on OR that was discussed in Section 3.4.4. For the purpose of the discussion in this section, the terms OP, CA and OR will be used interchangeably.

Research Question #2: Can the OR be measured pragmatically? Can the impact of KM on OR be measured pragmatically?	Objective: To determine if OR (as defined in this research) can be measured. Also, to determine if the impact of KM on OR can be measured and how previous attempts to make such measurements can inform this research. The findings are used in formulating the OR section of the questionnaire used in the research.	Methodological Approach: Conduct a literature review to determine if and/or how OR has been measured and how such measurement can be used in this research. Incorporate the most suitable form of the measurement of OR, within the context of this work, into this research.

Table 3.6.3.1: Research question #2

As mentioned in Section 3.3.4, an attempt to measure OR is represented by the work of Horne and Orr (1998) and their 74-item organizational resilience inventory assessment tool; the propositional study of Braes and Brooks (2010) also attempted to identify concepts that contribute to OR through the wide lens of various business domains and worldly viewpoints. The work of McCann et al. (2009), which examined companies' responses to environmental turbulence, did not consider KM as the primary independent variable (which is something that this research focuses on).

Nonetheless, the approaches to measuring OR presented in Section 3.3.4 proved to be informative for this research, primarily in terms of the selection and the design of its data collection instrument.

As pointed out in Section 3.4.4.3, there are two primary works to consider when searching for publications concerning the impact of KM on OR and attempts to measure OR solely in terms of KM (where KM is considered the independent variable). The first is the work of Choi et al. (2008) concerning the impact of KM on OR, and the second is the work of Wu et al. (2010), which investigates the impact of KM on OP as measured by DM.

The work of Choi et al. (2008) was found to be the only practical work on KM and OP. In this study, the authors analyzed 131 Korean firms and assessed the synergistic relationship between KM strategies and the impact of these strategies on OP.

The work of Wu et al. (2010) can be classified as work that practically examines the impact of KM on OR using DM. The lack of works that identify practical, quantitative ways of measuring the contributions of KM to OR was also noted by Wu et al. (2010, pg. 400).

Wu et al. (2010, pg. 400) seek to explore, using DM's Bayesian network algorithm and rough set theory, 'highly diverse KM patterns that distinguish lower and higher-performing companies.' The justification for their work comes from what they see as (2010, pg. 397) 'increasingly numerous concerns about whether the KM efforts can be fairly reflected and transformed into business performance.' As an indicator of performance, the authors chose return on assets (ROA). While the work of Wu et al. (2010) does not provide full details and also does not clearly identify what the authors refer to as the KM style that has the greatest effect on ROA, the text does identify certain hidden patterns (2010, pg. 401); the authors' analysis also showed that two of the most important factors affecting ROA (the dependent variable) are the 'purpose' and 'tacit' attributes (2010, pg. 399), which were not defined in their work. At the same time, the authors also state that one of the limitations of their research was the low number of attributes used in analysis – a factor that was adjusted for in this doctoral research based on the results of the literature review.

Thus, while the literature review did not identify any practical work that investigates the impact of KM on OR using DM, a number of theoretical works were identified. The identified works were discussed in Section 3.4.3.

3.6.4 Gaps Identified in the Literature Review and the Position of this Research Towards Addressing Them

The review of the literature revealed several gaps which are listed here, in order of their importance to this research; each gap is followed by a description of a method of addressing it.

Shortcoming # 1:

As concluded in Section 3.2.7, discussed in Section 3.4 and 3.6 and mentioned throughout the literature review chapter, there is a gap in the literature that addresses the impact of KM on OR (as defined in this work), from both the theoretical and practical perspectives. Because of the existing lack of literature, there is a need to refer to alternative methods of establishing the relationship between KM and OR. This research addresses the lack of literature by focusing on the impact of KM on OR through the following steps, which were discussed in the above-mentioned chapters:

- First, the argument is developed that, for the purpose of this research, OP/CA and OR are equivalent;
- Second, the KM process-based framework of Burnett et al. (2004; 2013) is mapped onto the McKenzie and van Winkelen (2004) six competence framework to provide KM-based questionnaire questions, which are grouped into KM competencies. This mapping is presented in Appendix IV;
- Third, the OR and OP/CA questionnaire questions were based on the literature review, forming KM-based questions as independent variables and OR as the dependent variable used in this research; and the questionnaire's OP-related questions, also derived from the literature review, serve the purpose of ensuring the statistical correlation between OP and OR (presented in Appendix III) to further justify the equation of OP with OR.

As the final result, the research addresses the impact of KM on OR.

Shortcoming # 2:

As stated by Corte-Real et al. (2014, pg. 176), 'there is little understanding of how BI&A systems may effectively be used and create positive impact on [an] organization.' The discussion in the work of Wu et al. (2010) also illustrates the lack of research in the area of applied quantitative studies of the impact of KM on OP (and, therefore, on OR) using DM (2010, pg. 400)

Through the use of practical DM models based on the literature review, this research demonstrates how DM can be methodically used to measure the impact of KM on OR. The outcomes of this research allow the following:

- Arriving at the resilience score (called the OR-Score in this research), which is derived by DM and based on replies to the questionnaires;
- Determining the KM processes and KM activities that affect OR (either positively or negatively) and the extent to which they do so;
- Comparing the KM activities and processes of resilient and non-resilient organizations to determine which KM activities are responsible for either a low or high OR–Score, and to what extent;
- Inspecting which KM activities and KM processes are related to each other; and
- Determining the level of accuracy of the resultant DM model used to measure the impact of KM on OR.

The outcome of the research is the first comprehensive examination of how DM can be used as a measurement instrument to measure the impact of KM on OR.

Shortcoming # 3:

No published works were found that attempted to map KM processes, as presented by Burnett et al. (2004; 2013), onto the McKenzie and van Winkelen model (2004); such works may have been used to ensure that no KM processes have been omitted from consideration when constructing the measuring instrument and the DM models.

For the needs of this research, the mapping between the KM processes of Burnett et al. (2004; 2013) was established and presented in Chapter 4.

3.6.5 Summary

This section summarized the literature review that was conducted for the purpose of this thesis. The section focused on answering RQs # 1 and #2, as those two research questions (as stated in Chapter 4, which describes the methodology used in this research) were derived from the literature review. In addition, the gaps in the literature were re-stated and solutions to the shortcomings identified by this research were provided.

The conclusion of the literature review is that DM is an excellent tool for addressing the measurement of the impact of KM on OR. The superiority of DM over the classical statistical tools is illustrated in the following chapter.

CHAPTER FOUR: METHODOLOGY

4.1 Introduction

The purpose of this chapter is to justify the research methods selected and employed within this research, to explain how these methods have been applied in relation to this work and to provide the overall map of and execution steps for this study.

This chapter begins with an introduction to the main research paradigms and then moves on to a discussion of the stages of the research, followed by a detailed description and justification of the research design. The chapter ends with the research plan conceptualized by the researcher, which considers some of the key aspects of the research, including the philosophical perspective, the applicable research type, the research approach, the research time horizon and the research methods. The presentation of the topics in this chapter follows the research process presented in Fig. 4.2.1. The overall research process is presented next.

4.2 Research Process

Saunders et al. (2009, pg. 5) define research 'as something that people undertake in order to find out things in a systematic way, thereby increasing their knowledge.' Saunders et al. (2009, pg. 5), as well as Rajasekar et al. (2006, pg.1), stress the importance of the 'systematic way' in finding solutions to research questions and problems. Rajasekar et al. (2009, pg.1) presents six main objectives of research:

1. To discover new facts;
2. To verify and test important facts;
3. To identify cause and effect relationships;
4. To develop new concepts, theories and scientific tools that could be used in solving scientific and nonscientific problems;
5. To find solutions to social, scientific and nonscientific problems; and
6. To solve or overcome everyday life problems.

The research process used in this study is that proposed by Saunders et al. (2009, pg. 11), which consists of several steps that, when completed, satisfy the researcher's aims and objectives. The research process proposed by Saunders et al., and modified for the purpose of this study, is presented in Fig. 4.2.1 and described in the following sections of this chapter.

The processes presented in Fig. 4.2.1 are discussed next, in the order shown.

4.2.1 Strategies of Inquiry

As stated by Rajasekar et al. (2006, pg. 1), 'research is a logical and systematic search for new and useful information on a particular topic. It is an investigation of finding solutions to scientific and social problems through objective and systematic analysis.'

Research is very frequently divided into two main paradigms: the qualitative and quantitative paradigms, where, as stated by Krauss (2005, p. 759), 'a paradigm can be defined as the basic belief system or world view that guides the investigation.' There is, however, a slight twist to the 'qualitative/quantitative paradigms' because, as stated by Creswell (2003, pg. 4), 'mixed methods research has come of age.' The mixed method approach is briefly introduced at the end of this section.

Quantitative research, according to Rajasekar et al. (2006, pg. 4), 'is based on the measurement of quantity or amount. Qualitative research is concerned with qualitative phenomenon involving quality. It is non-numerical, descriptive, applies reasoning and uses words. It aims is to get the meaning, feeling and describe the situation.' In more common terms, qualitative research seeks to consider the context of the situation in analysis, whereas the quantitative research is mainly context-independent.

Krauss (2005, pg. 767) goes to great lengths to illustrate key differences between the epistemologies of qualitative (naturalist/constructivist) and quantitative (positivist) research paradigms by portraying their differences as reflecting ontological views regarding the nature of reality: 'In the positivist paradigm, the object of study is independent of researchers; knowledge is discovered and verified through direct observations or measurements of

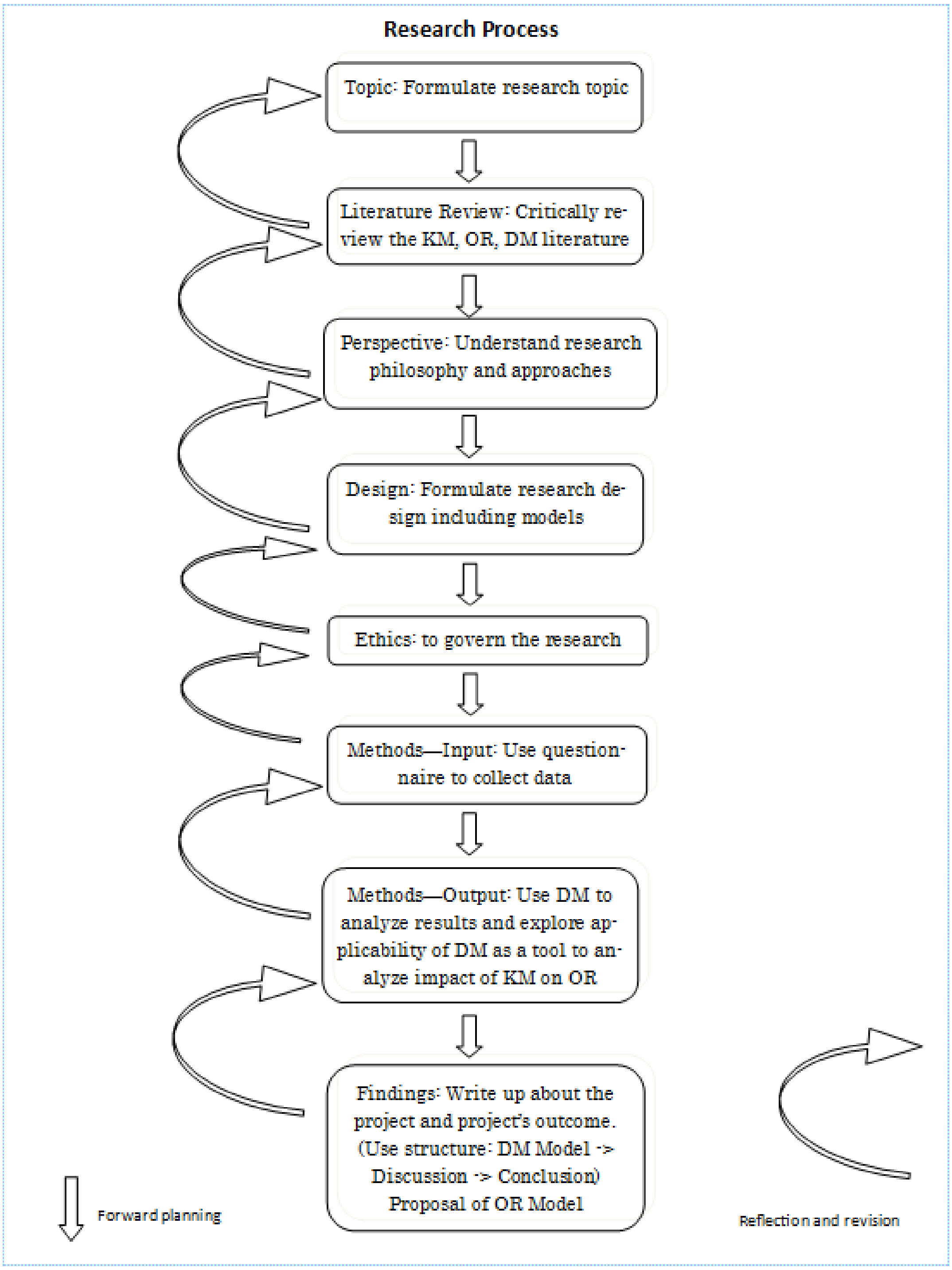

Research Process
Topic: Formulate research topic
Literature Review: Critically review the KM, OR, DM literature
Perspective: Understand research philosophy and approaches
Design: Formulate research design including models
Ethics: to govern the research
Methods—Input: Use questionnaire to collect data
Methods—Output: Use DM to analyze results and explore applicability of DM as a tool to analyze impact of KM on OR
Findings: Write up about the project and project's outcome. (Use structure: DM Model -> Discussion -> Conclusion) Proposal of OR Model
Forward planning
Reflection and revision

Fig. 4.2.1: The research process utilized in this study

phenomena; facts are established by taking apart a phenomenon to examine its component parts. An alternative view, the naturalist or constructivist view, is that knowledge is established through the meanings attached to the phenomena studied; researchers interact with the subjects of study to obtain data; inquiry changes both researcher and the subject; and knowledge is context and time dependent' (Krauss, 2005, pg. 759).

Another interesting point made by Krauss (2005, p. 760) is that qualitative research is based on relativistic, constructivist ontology that states that reality is not objective; rather, there are multiple realities shaped by human experience of the phenomenon of interest. From the positivist view, the world is much more ordered and deterministic and can be controlled and predicted by the laws of cause and effect and observations.

In addition to the classification of research paradigms, one can also classify research based on its type. According to Rajasekar et al., (2006, pg. 3) research is broadly classified into two main classes: 1) fundamental or basic research, or, 2), applied research.

'Basic research is an investigation on basic principles and reasons for occurrence of a particular event or process or phenomenon. It is also called theoretical research' (Rajasekar et al., 2006, p. 3). 'In applied research one solves certain problems employing well known and accepted theories and principles. A research, the outcome of which has immediate application is also termed as applied research' (Rajasekar et al., 2006, pg. 4).

Furthermore, in addition to the classification of research into basic and applied types, other classifications are often presented by practitioners and educators. Walliman (2011, pg. 7) recognizes research classification based on the objectives of the research. Such an approach to classification, based on Walliman's work, along with comments regarding its applicability to this research, is presented in Table 4.2.1, below.

Classification name:	**Characteristics of classification:**	**Position of classification in relation to this research:**
Categorization	Involves forming a typology of objects, events or concepts that can later be useful in	At this phase of the study this approach is not applicable. It is conceivable that this

	explaining, among other things, what ‘elements’ belong together and how.	approach could be employed when looking for association between independent variables, among other things, in follow up studies.
Explanation	Seeks to explain phenomena that are not, or are only partially, understood.	Not applicable to this research; however, it is feasible that a study such as this, given a sufficient amount of data, could seek to explain which KM processes are the most important for achieving OR, for example.
Prediction	Commonly made on the basis of an explanation of a phenomenon in anticipation of future events, associations, inner workings and causation.	This approach could be very well utilized in another study that could not only validate this study but also provide actionable insights regarding the impact of KM processes on OR.
Understanding (making sense of)	Seeks to provide a complete explanation of a phenomenon, including the explanation of why and how things happen.	This approach could be feasible in follow-up studies, provided meaningful and accurate results can be obtained from DM algorithms.
Control	Attempts to find a way to control a phenomenon.	This approach is not applicable to the current research, but it is conceivable that this

		type of research could be conducted following the 'understanding' study type.
Evaluation	Makes judgments, in absolute terms or on a comparative basis, about the quality of objects or events.	Not a focus of this research as this work does not seek to evaluate various KM initiatives and their impact on OR; however, this type of work cannot be excluded as a possibility for future research.

Table 4.2.1: Classification of the research based on objectives

Research classification based on research type can be also extended further, beyond the segregation of research into basic and applied research and beyond segregation based on research objectives. Walliman (2011, pg. 9) identifies ten major research types: action, historical, comparative, descriptive, correlation, experimental, evaluation, ethnogenic, feminist and cultural. In addition, Saunders et al. (2009, pgs. 587, 592-593) provide two additional classifications: explanatory and exploratory research.

The applied research type reflects the nature of this work, as the research problem is well-defined and the findings have practical relevance. Applied research is defined by Saunders et al., (2009, pgs. 587, 592-593) as '[r]esearch of direct and immediate relevance to practitioners that addresses issues they see as important and is presented in ways they can understand and act upon,' which agrees with the previously provided definition of applied research offered by Rajasekar et al., (2006, pg. 4).

4.3 Formulation of Research Topic and Research Questions

In order to illustrate the process of selecting the research topic and research question, the following supporting material is presented next.

4.3.1 The Research Problem

Saunders (2009, pg. 25) suggests two ways of considering the research idea: following either rational or creative thinking. As part of the rational approach, Saunders lists the following items to consider: examining one's own strengths and interests, looking at past project titles, discussion, searching the literature and scanning the media. The following elements inform the creative thinking position: keeping a notebook of ideas, exploring personal preferences, relevance trees and brainstorming. Within the context of the work of Saunders, the research idea for this study was primarily developed out of the personal preference of the researcher, which manifested itself as curiosity as to why some organizations perform well regardless of business conditions. The second aspect was the researcher's professional experience with DM. According to Creswell (2003, pg. 22), '[i]nto the mix of choice also comes the researcher's own personal training and experiences. An individual trained in technical, scientific writing, statistics, and computer statistical programs who is also familiar with quantitative journals in the library would most likely choose the quantitative design.' So, while the choice of research design is discussed later in this chapter, the researcher's professional experience, along with personal curiosity, naturally translated itself into a scientific research project which asks, in more casual terms, given the advances in technology, especially in the area of DM, can 21st century DM tools be used to uncover the intricate relationships that may exist between KM and OR?

The overarching guide for the generation of the research questions was the issue of generating new insights while keeping the answers specific, measurable and achievable, in line with suggestions of Saunders (2009, pg. 35). The research questions used in this work were introduced in Chapter 1 and are also presented later in this chapter.

Five specific questions were solidified based on the findings of the literature review. Research questions #1 and #2 relate to the concepts of DM, KM, OR and the relationships between them and are answered by the literature review. Research questions #3 to #5 are more applied in nature, and their answers arise from the DM component of this research.

The review of the literature can be grouped into three logical parts: KM-related, OR-related and a section that encompasses KM, OR and the impact of KM on OR through a DM lens. The literature review presented in Chapter 3 is structured according to the Fig. 3.1.1, presented in Chapter 3.

4.4 Literature Search & Review

Preceding the research, a literature search and review were conducted. The literature search was conducted, on a regular basis, throughout the duration of the research project. Initially, given the multi-disciplinary nature of this research, the search focused on theoretical models that could be adapted for the purpose of this research and later moved on to the identification of prior work that could provide a starting point for this research.

The overall structure of the literature search and review can be classified into four areas: 1) knowledge management, 2) organizational resilience, 3) data mining, and 4) application of DM with relation to KM and/or OR (encompassing the previous three searches and reviews in the context of DM).

4.4.1 Topics

The literature search initially sought to identify existing works that dealt with KM, OR and DM used in the business/social context. This included inspecting various perspectives on KM and seeking the definition of OR that was most acceptable to the author of this research, as well as determining its associations with organizational performance and competitive advantage. Once the literature search and review had been largely completed with regard to KM and OR, it was limited to occasional searches in order to determine the presence of any new material that could be used in this research; the primary focus of the search and review process then came to encompass KM and OR data mining. With regard to DM, several searches were conducted. Some of the searches focused on the DM algorithms themselves and their applicability to the subjects of this research. Other searches focused on the existing use of DM in business, with a special focus on its practical applications and the issues encountered and conclusions arrived at in such studies.

The overall feel as an outcome of these searches is that the KM area possesses large body of knowledge in terms of various perspectives, theory and application going back to the work of Polanyi (1966).

The area of resilience appears to have a large body of knowledge related to personal and organizational resilience defined slightly differently to be of interest to this research, with a large body of knowledge regarding OR as a concept of recovery after some catastrophic (to the organization) event.

The number of authors who address the impact of KM on OR is very limited; principally, the work of McCann et al. (2009) is highly applicable to this research. (There were other key works that formed the 'foundation' for this research; these have been mentioned in Chapter 3.)

With DM receiving a great deal of attention in the last few years in the business world, the number of DM-based (or BI-based) publications has increased significantly over the last decade. Yet, to date, the number of works that focus on measuring the successful use of DM in the real world is not large, and some writers even question DM's role as a success factor for organizations. Yet, a number of works, discussed in Chapter 3, provided an excellent background to this research.

In terms of the topics searched for, the following broad categories were considered:

- Knowledge, KM, knowledge economy, knowledge processes, knowledge value, knowledge value creation;
- Organizational resilience, competitive advantage, business performance, strategic positioning, profitability, survivability, competitiveness, greatness, efficiency, effectiveness;
- Data mining, BI, machine learning, learning algorithms, predictive analytics;
- Social science research methodologies, research methods; and
- Research philosophies, research approaches, research strategies.

4.4.2 Sources

In relation to the topics identified above, the following table outlines some of the keywords which were identified and used within the literature search and review:

Topic:	Keywords:
Knowledge, KM, knowledge economy, knowledge processes.	Knowledge, KM, knowledge economy, organizational learning, knowledge perspective, knowledge processes, post-industrial society, information society, knowledge-intensive.
Organizational resilience, competitive advantage, business performance, strategic positioning, profitability, survivability, competitiveness, greatness.	Organizational resilience,, resilience, competitive advantage, survivability, adversity, organizational performance, efficiency, effectiveness.
Data mining, business intelligence, machine learning, learning algorithms.	Data mining, business intelligence, big data, impact of data mining, algorithms, practical data mining, machine learning, data science, predictive analytics.
Social science research methodologies, research methods.	Social science, research, research methodology, research methods, quantitative research, qualitative research, mixed methods research, analysis of data, data input.
Research philosophies, research approaches, research strategies.	Research philosophy, epistemology, ontology, axiology, positivism, post-positivism, research strategy, research techniques, philosophical underpinning.

Table 4.4.2.1: Main keywords used in literature searches

A number of databases were used as key sources for relevant journals, journal articles and bibliographies used in this research:

- Business Source Partner
- Emerald
- LISTA (Library, Information Science and Technology Abstracts)
- SAGE Journal Online
- ScienceDirect
- Social Science Citations Index
- Web of Knowledge

The following journals were regularly scanned for relevant articles using the databases listed above (however, research was not necessarily limited to the following list):

- Harvard Business Review
- Expert Systems with Applications
- Decision Support Systems
- Applied Mathematics and Computation
- Neurocomputing
- The Journal of Knowledge Management
- The Journal of Information Science
- The Journal of Knowledge and Process Management
- Journal of Business Strategy

4.5 Philosophical Assumptions

According to Creswell (2003, pg. 6), 'stating a knowledge claim means that researchers start with certain assumptions about how they will learn and what they will learn during their inquiry.' There are certainly no shortages of terms used in the academic and non-academic publications that refer to philosophical assumptions. According to sources cited by Creswell (2003, pg. 6), the most common terms used in reference to philosophical assumptions are paradigms, ontologies and research methodologies. The four schools of thought regarding knowledge claims presented by Creswell are post-positivism, constructivism, advocacy/participatory, and pragmatism.

Post-positivism – Postpositive knowledge claims have traditionally governed claims about what warrants knowledge. Additional terms referred to in this view include the scientific method, quantitative research, positivist/post-positivist research, empirical science and post-positivism. As presented by Creswell (2003, pg. 7), the term post-positivism reflects the fact that simply maintaining a positivist view no longer suffices; this challenges the notion of the absolute truth of knowledge, recognizing that the notion of 'absolute truth' may not be appropriate, especially when studying the actions and behaviors of humans.

Post-positivism reflects a deterministic philosophy in which causes probably determine outcomes, making the studies of problems in which causes influence outcomes highly applicable to this school of thought. The knowledge that is developed using the post-positivist approach is based on the observation and measurement of the objective reality that is thought to exist in the world. Developing numeric measures of observations and studying the behavior of individuals have become the key approaches for a post-positivist.

The post-positivism position originally presented by Phillips and Burbules (2000) is cited by Creswell (2003, pg. 7) as having five key assumptions:

1. Knowledge is conjectural (and anti-foundational) – absolute truth can never be found;
2. Research is the process of making claims and then refining or abandoning some of them for other claims that are more strongly warranted;
3. Data, evidence and rational considerations shape knowledge;
4. Research seeks to develop relevant true statements that can serve to explain the situation that is of concern or that describes the casual relationship of interest; and
5. Being objective is an essential aspect of competent inquiry, and, for this reason, researchers must examine methods and conclusions for bias. (Reliability and validity are considered extremely important in quantitative research.)

With the post-positivist view most closely matching the view of the author of this research, the other perspectives are mentioned below only to ensure the completeness of the discussion of these perspectives on knowledge claims. The justification for selection is presented in the next section.

Constructivism –Socially constructed knowledge claims hold the assumption that individuals seek an understanding of the world in which they live and work. They develop subjective meanings from their experiences – meanings that are directed towards certain objects or things (Creswell, 2003, pg. 8). As they are affected by personal experiences, these meanings vary greatly, leading the researcher to look for a number of (complex) views. The goal of the research, then, is to rely as much as possible on participants' views of the situation under investigation.

Advocacy/Participatory – The advocacy/participatory view is relatively new, as it arose during the 1980s and 1990s from individuals who felt that the lack of theories and laws that were an appropriate fit for marginalized individuals or groups or did not properly address issues of social justice. The advocates of this position believe that inquiry needs to be integrated with politics and a political agenda, implying that research should contain an action agenda for reform that would positively affect the lives of research participants and/or the organizations in which individuals work, as well as the researcher's life.

Pragmatism – (Creswell, 2003, pg. 11) There are many forms of pragmatism. For a number of them, knowledge claims arise out of actions, situations, and consequences instead of antecedent conditions (as was the case in post-positivism). Applicability, or "what works," and the solution to a problem are very important to this perspective. The emphasis is on the problem, rather than on the method, so all approaches are acceptable if they indeed help to understand the problem.

4.5.1 Selection of the Research Approach

The discussion of the research, the research processes and research design takes place in the context of the model presented by Creswell (2003, p.5).

From the philosophical perspective and taking into account assumptions about what constitutes knowledge claims, the post-positivist approach best fits this study. The choice of the post-positivist knowledge claim, rather than simply the positivist claim, is made on the basis that, just as maintaining a positivist view that focuses on the notion of absolute truth no longer suffices when studying actions and behaviors of humans, the study of organizations (which employ humans and are therefore a form of social group) also needs to challenge the notion of absolute truth.

The deterministic nature of the post-positivist philosophy generally fits well with the views and beliefs of the author of this work in that, for the most part, observation and measurement of objective reality can, to a significant extent, provide an accurate 'view of the world'. As was stated by Creswell (2003, pg.6), '[d]eveloping numeric measures of observations and studying the behavior of individuals become the key approach for a post-positivist.' The organization is viewed as a social and living entity made up of one or more individuals; thus, organizations tend to be well suited to the post-positivist approach. It is expected, therefore, that this work will produce relevant, objectively true statements about a new way of measuring of the impact of KM processes on OR.

Clearly, because this study involves organizations (which are viewed as living entities), there are certain elements of the constructivist knowledge claims that seem appealing. Based on the fact that such claims are constructed by the individual's meaning of the world and focus on the impact of social and past events on shaping such views, it can be argued that these claims are applicable to organizations. What makes a positivistic approach more appealing, however, is the desire for the individual person's view and independent experience in testing the use of DM as a tool for measuring the impact of KM on OR and the identification of key processes.

From the general procedures of research, or what Creswell calls the 'strategies of inquiry (2009, pg. 5)' perspective, several choices made under this category are discussed. From the comparison of basic research (BR) and applied research (AR) and the discussion in Chapter 4.2.1, it appears that applied research is the most appropriate classification for this work and it is therefore classified as such.

Note that the focus of this research has changed slightly over time. The initial goal of the study was the actual measurement of the impact of KM processes on organizations, using data mining for analysis and a questionnaire as the data collection instrument. Due to the very low questionnaire return rate, however, the study could not be completed, which forced a change in the focus of the research.

The very low actual return rate (around 1%) to the questionnaire instrument used in an attempt to measure the impact of KM on OR made measurement impossible, as the number of replies was too small to draw conclusions from. Because of the low questionnaire response rate, the research has therefore been altered, with a new focus on applied research testing the suitability of DM tools for evaluating the impact of KM on OR.

4.6 Main Study: Research Design

According to Saunders et al. (2009, pg. 136), research design is a general plan of how one intends to go about answering the research question(s). According to Saunders et al. (2009, pg. 137) the design should contain the following elements:

- Clear objectives derived from the research questions;
- Specification of sources for the data;
- Consideration of constraints affecting the design; and
- A discussion of ethical issues that affect the research.

The aspects of research design mentioned above are discussed in the next section, and the overall research design structure is presented in Fig. 4.6.1.

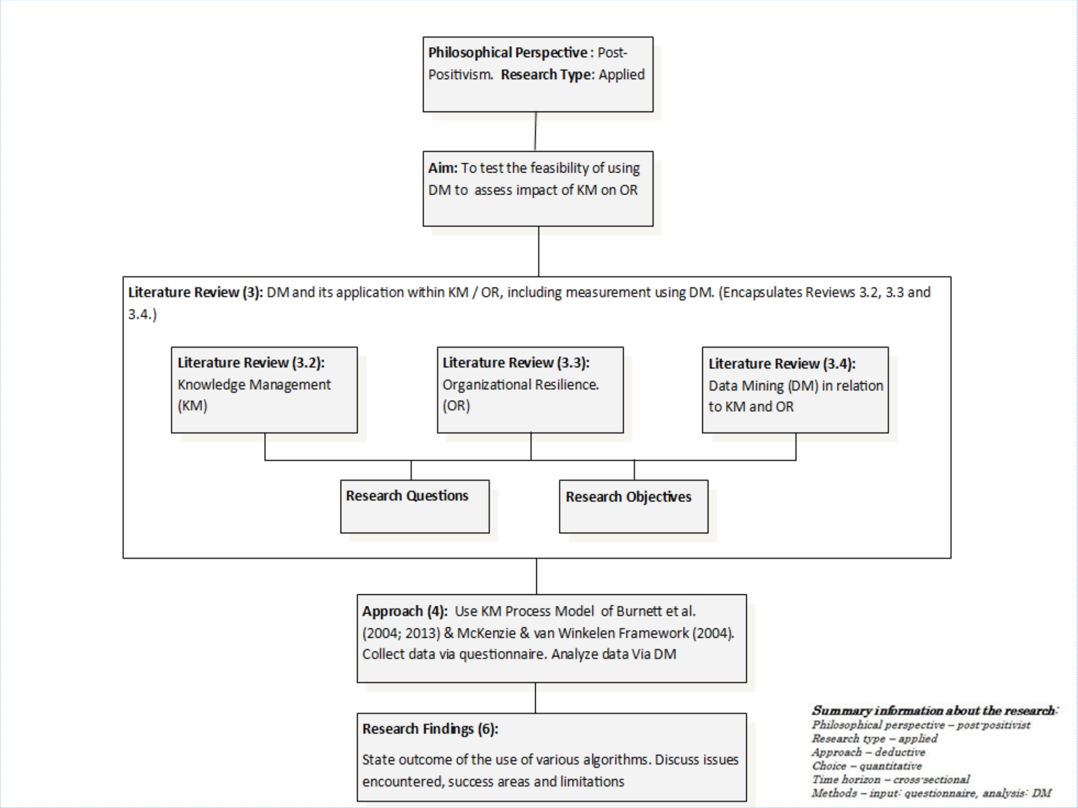

Philosophical Perspective: Post-Positivism. Research Type: Applied
Aim: To test the feasibility of using DM to assess impact of KM on OR
Literature Review (3): DM and its application within KM / OR, including measurement using DM. (Encapsulates Reviews 3.2, 3.3 and 3.4.)
Literature Review (3.2): Knowledge Management (KM)
Literature Review (3.3): Organizational Resilience. (OR)
Literature Review (3.4): Data Mining (DM) in relation to KM and OR
Research Questions
Research Objectives
Approach (4): Use KM Process Model of Burnett et al. (2004; 2013) & McKenzie & van Winkelen Framework (2004). Collect data via questionnaire. Analyze data Via DM
Research Findings (6): State outcome of the use of various algorithms. Discuss issues encountered, success areas and limitations
Summary information about the research:
Philosophical perspective – post-positivist
Research type – applied
Approach – deductive
Choice – quantitative
Time horizon – cross-sectional
Methods – input: questionnaire, analysis: DM

Fig. 4.6.1: Research design structure

4.6.1 Research Planning

'The purpose of the research plan is to take the initial research problem and decide how it will be researched' (Walliman, 2011, pg. 40).

While Fig. 4.6.1 specifies the research design structure, and Chapter 1 listed the aims and objectives of this research, the methodological approaches utilized in this research are listed in Table 4.6.1.1, below:

Aim of research: **To test the feasibility of using DM to assess the relationship between and impact of KM and OR.**		
Research Questions:	Objectives:	Methodological Approaches:
Research Question #1: What prior research exists regarding the application of DM with respect to KM and OR	Objective: To determine the feasibility of using DM when evaluating KM, OR and/or the impact of	Methodological Approach: The first round of the literature review focused on the fields of KM and

and the impact of KM on OR and what are the known relationships between KM and OR?	KM on OR. Also, to determine the applications of DM techniques that have been developed in support of KM and OR as well as to identify the areas of convergence between DM, KM and OR.	OR. Then, a second round of the literature review focused on examining the impact of KM on OR through a DM lens. (Overall, the literature reviews employ a circular approach.) Mapping of theoretical and practice-based research in DM, KM and OR from the literature review.
Research Question #2: Can OR be measured pragmatically? Can the impact of KM on OR be measured pragmatically?	Objective: To determine if OR (as defined in this research) can be measured. Also, to determine if the impact of KM on OR can be measured and how previous attempts to make such measurements can inform this research. The findings are used in formulating the OR section of the questionnaire used in the research.	Methodological Approach: Conduct a literature review to determine if and/or how OR has been measured and how such measurement can be used in this research. Incorporate the most suitable form of measuring OR, within the context of this work, into this research.
Research Question #3: Which KM processes are the most influential for OR?	Objective: To explore the use of DM in order to test the suitability of applying	Methodological Approach: Create data collection instrument, administer

	DM to the primary grouped data composed of the questionnaire answers, to assess their relationship with OR.	it and collect replies. Use the DM, as an analytical tool used in arriving with the answer to the research question #3.
Research Question #4: Can a methodological approach be developed to examine the relationships between KM and OR, utilizing DM?	Objective: To develop and apply a DM-based methodological approach for the analysis of data gathered from the use of the questionnaire and the generation of valid findings for this research.	Methodological Approach: Develop an analytical and practical approach through a synthesis of BI, KM and OR.
Research Question #5: Which are some of the main challenges when employing DM for the purpose of determining the impact of KM on OR?	Objective: To identify the main issues (data, algorithm, error, algorithm parameters) associated with the use of DM for the purpose of measuring the impact of KM on OR.	Methodological Approach: Investigate the challenges and requirements associated with each DM algorithm utilized in this research.

Table 4.6.1.1: Aim, objectives, research questions and methodological approaches

The research was set out to include the following main aspects:

- Literature search and review – [Section 4.4].
- KM process model selection – [Section 4.6.2].
- Data collection instrument (questionnaire) construction and validation – [Section 4.8.8].
- Selection of recipients of questionnaire – [Section 4.8.6].
- Time horizon selection – [Section 4.8.2].

- Questionnaire pilot testing – [Section 4.8.7].
- Questionnaire administration – [Section 4.8.6].
- Data manipulation (into the SQL Server) – [Section 5.4].
- DM model construction/use/evaluation – [Chapter 6].

Each one of key research aspects mentioned above (belonging to this chapter) is addressed and justified in this chapter.

4.6.2 KM Process Model Selection

This section of work builds on the definitions introduced in Chapter 2 and in the KM-based literature review in Section 3.2.

The work of Alavi and Leidner (2001) in the area of KM leads them to note that KM is largely looked at from a process-based perspective that involves various activities. Numerous researchers, including Wiig (1993), DiBella and Nevis (1998), Liebowitz (1999), Alavi and Leidner (2001) and Burnett et al. (2004; 2013), have proposed process-based KM models that differ in the number of processes used. Alavi and Leidner (2001, pg. 114), writing about KM processes/activities, state that '[s]light discrepancies in the delineation of the processes appear in the literature, namely in terms of the number and labeling of processes rather than the underlying concepts.' The purpose and design of such KM processes, as stated by Fink and Ploder (2007, pg. 705), are such that organizational profitability and competitive advantage in the marketplace are improved.

This research builds on the process-based view of the firm, using the process-based KM model adapted from Burnett et al. (2004, pg. 29) and further expanded upon with the McKenzie and van Winkelen (2004) model. (The Burnett et al. model is presented in Fig. 3.2.3.1.1, Section 3.2.3.1.)

Burnett et al. (2004, pg. 10) define the KM processes as follows:

'• Acquisition and Learning – learning, acquiring new knowledge from people, books, websites etc.

'• Storage and Maintenance – storing knowledge to make it easily accessible to all who may require it and ensuring that it is kept up-to-date and relevant.

'• Application and Exploitation – putting knowledge to use, deriving benefit from it in carrying out work.

'• Dissemination and Transfer – proactively sharing knowledge with others (formally or informally) on a one-to-one or a one-to-many basis verbally, in written form, electronically etc.

'• Knowledge Creation – using knowledge to create value through new ways of doing things, new products or services.

'• Performance Measurement – determining how well the above activities are carried out and how they impact on work focusing on measurable benefits.'

While the model presented by Burnett et al. (2004, 2013) tends to confirm the findings of Alavi and Leidner (2001, pg. 114), it has been selected as the KM process model because it includes all of the major KM-related processes that are referred to in the KM literature review as necessary in order for an organization to gain competitive advantage and to improve its well-being – making the inclusion of the 'application and exploitation' process a very important part of the overall model. Moreover, the Burnett model clearly shows the connections between each KM process and, in addition to including the application and exploitation process, it views the knowledge creation process as its centerpiece. This view is in line with the stance taken in this research that, in addition to the creation of operational/business knowledge, it is also critical to create (as well as later to act upon) knowledge concerning surrounding business conditions. Such environmental scanning and sense-making appear to be the key prerequisites for organizational resilience (iJet International Inc., 2008, pg. 5; McCann et al. 2009, pg. 45; Hamel & Valinkangas 2003, pg. 3; Sundstrom & Hollnagel, 2006, pg. 9).

The expansion of the model comes from the work of McKenzie and van Winkelen (2004), whose general concept is illustrated in Fig. 4.6.2.1. The mapping between the KM-process-based model and the model of McKenzie and van Winkelen is presented in Fig. 4.6.2.2.

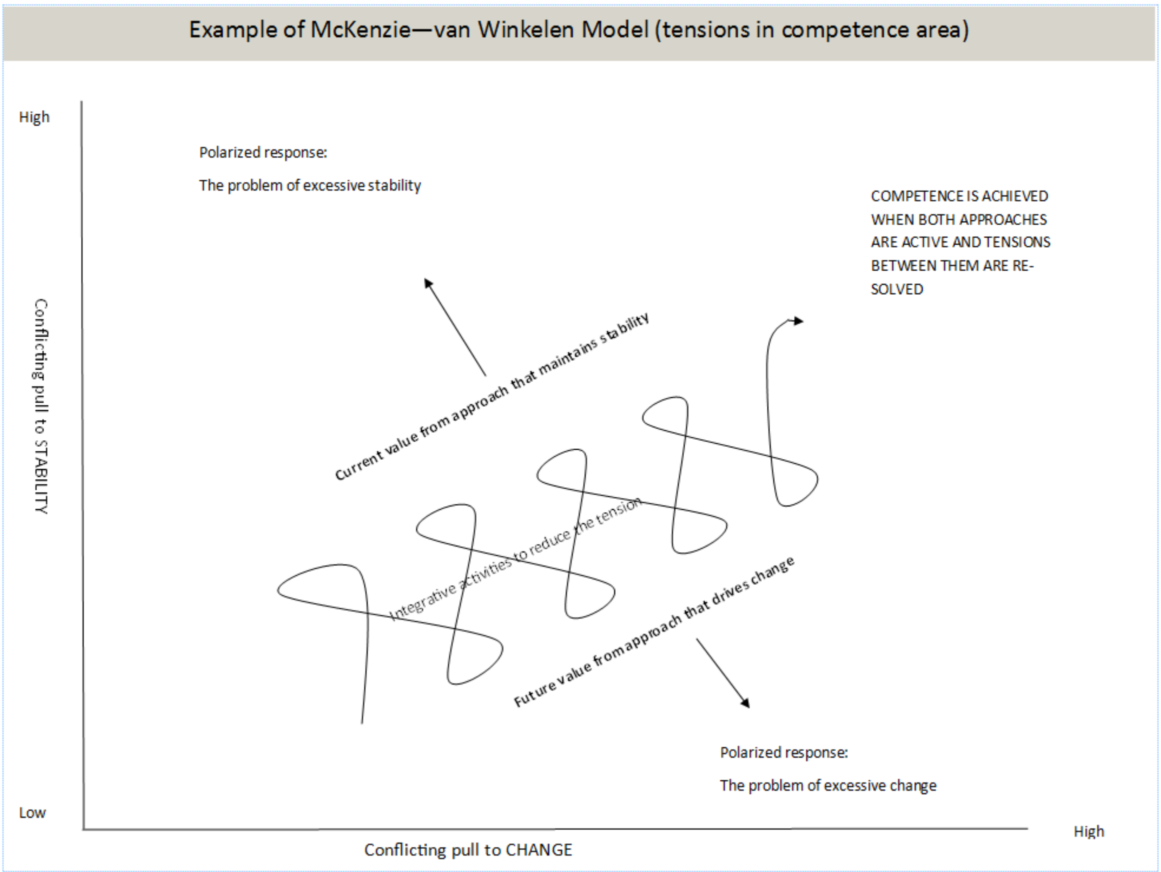

Example of McKenzie—van Winkelen Model (tensions in competence area)
High
Polarized response:
The problem of excessive stability
COMPETENCE IS ACHIEVED WHEN BOTH APPROACHES ARE ACTIVE AND TENSIONS BETWEEN THEM ARE RE-SOLVED
Conflicting pull to STABILITY
Current value from approach that maintains stability
Integrative activities to reduce the tension
Future value from approach that drives change
Polarized response:
The problem of excessive change
Low
Conflicting pull to CHANGE
High

Fig. 4.6.2.1: Developing each area of competence by resolving the tensions between approaches that maintain stability and drive change. [Derived from McKenzie and van Winkelen (2004, pg. 3).]

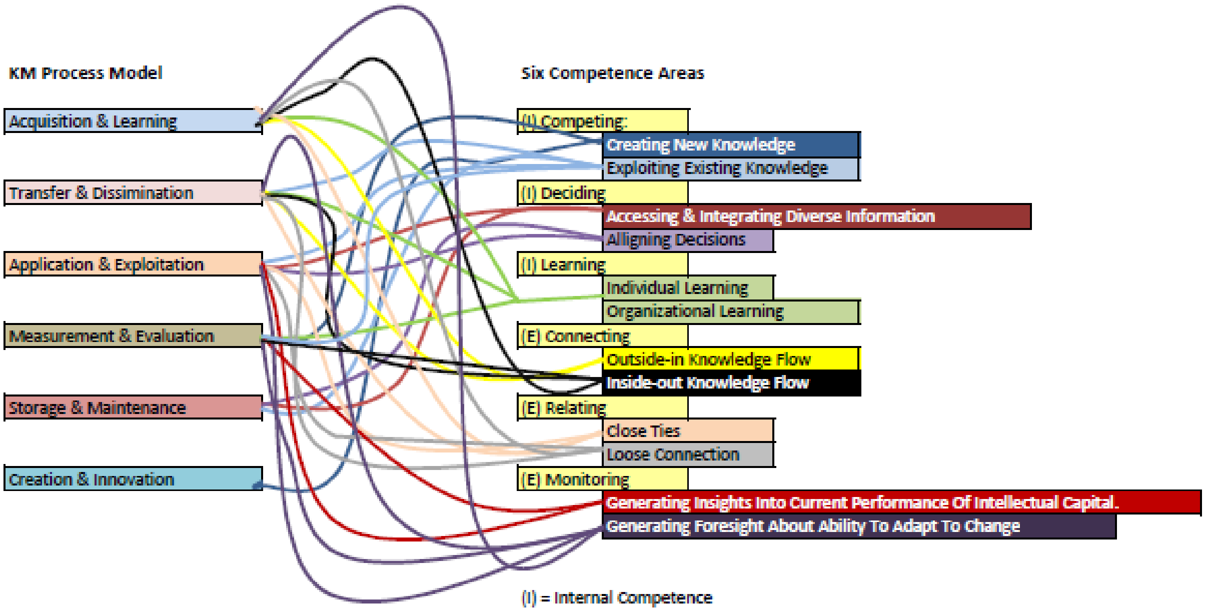
McKenzie and van Winkelen (2004) propose a model for leveraging the knowledge resources contained within an organization as well as for the improvement of operational effectiveness within the knowledge economy. The process-based model proposed by McKenzie and van Winkelen utilizes six competence areas (competing, deciding, learning, connecting, relating and monitoring) that are divided into two categories: those that are internal to an organization (encompassing the first three areas of competence) and those that are external to an organization (composed out of the last three areas of competence). The uniqueness of the model, which makes it greatly appealing as the preferred model for this research, is the fact that it considers, within each competence area, two opposing forces that act upon the competence area, creating a tension. One force attempts to utilize and maximize returns/value from current knowledge (therefore, an organization does not abandon its existing goals) and the other force pulls towards change, emphasizing the future and the future value to be derived from knowledge. Competence is obtained when both forces act and the tension is stabilized. Worth noting is the realization that paying too much attention to either of the force produces a polarized response that is detrimental to an organization (McKenzie & van Winkelen, 2004, pg. 3). Similarly to that of many other writers, the work of McKenzie and van Winkelen focuses on people, their interactions and environment.

One important point to note is the fact that 'classic KM-process based models', such as that adopted from Burnett et al., do not make an explicit distinction between the need to focus on both maximizing the benefits of existing KM processes and thinking forward and planning for the future. Robb (2000, pg. 27), emphasizes this point, noting the importance of both planning for the future and maximizing the existing opportunities and stating that "[a] resilient organization is able to sustain competitive advantage over time through its capability to do two things simultaneously:

- 'Deliver excellent performance against current goals, therefore maximizing current opportunities.
- 'Effectively innovate and adapt to rapid, turbulent changes in markets and technologies, therefore preparing for the future.'

The tension forces present in the McKenzie and van Winkelen model fill this exact gap, as the forces in all six competence areas are the balance between the current state of 'things' and the state of 'things' to come. Presented below, in Fig. 4.6.2.2, is a mapping that identifies one possible correlation of the six KM processes present in the Burnett et al. model with the model proposed by McKenzie and van Winkelen. This mapping exercise has been performed in order to ensure that each element of the Burnett et al. (2004; 2013) model maps onto at least one competence area and that each competence area is associated with at least one KM process. The mapping, especially the right-hand side (the competence aspect) was used as the primary guide for arriving at the questionnaire questions.

The Mapping of Most Common Influences of KM Processes on the Six Competence Areas and Vice Versa



(I) = Internal Competence
(E) = External Competence
Learning - Combined Categories Due To Extensive Similarities

Examples of functional/strategic relationship and outcome similarities between the concepts:

KM Processes: — Six Competence Area Processes

Acquisition & Learning — Learning

Individual Learning

Formal lectures, on-the-job learning, reflection, mentoring, coaching, etc.

Organizational Learning

Somewhat similar to the individual learning with greater emphasis given to collective learning (involving connecting) through dialogue, stories, metaphores, CoPs.

Connecting

Outside-In Knowledge Flow

Environmental scanning, analyzing scans and identifying opportunities and threats.

Inside-Out Knowledge Flow

Learning from the result of releasing information to the outside environment.

Relating

Close Ties

Allows for learning/sharing of experiances among networked members while protecting against the competition. (Exchange of tacit and explicit knowledge.)

Loose Connections

Similar to the 'close ties' functions but emphasis on codified, easily diffusable knowledge. Possible to learn through exteernal consultants, outsourcing.

Transfer & Dissimination — Competing

Exploiting Existing Knowledge

Provision of mechanisms allowing for the current knowledge to flow, in timely manner.

Learning

Individual Learning

Enable existing knowledge to flow for the purpose of learning. Learning through interaction with others, reflection.

Organizational Learning

Dissimination of knowledge through Intranet, email, etc. Knowledge databases. Lessosns learned.

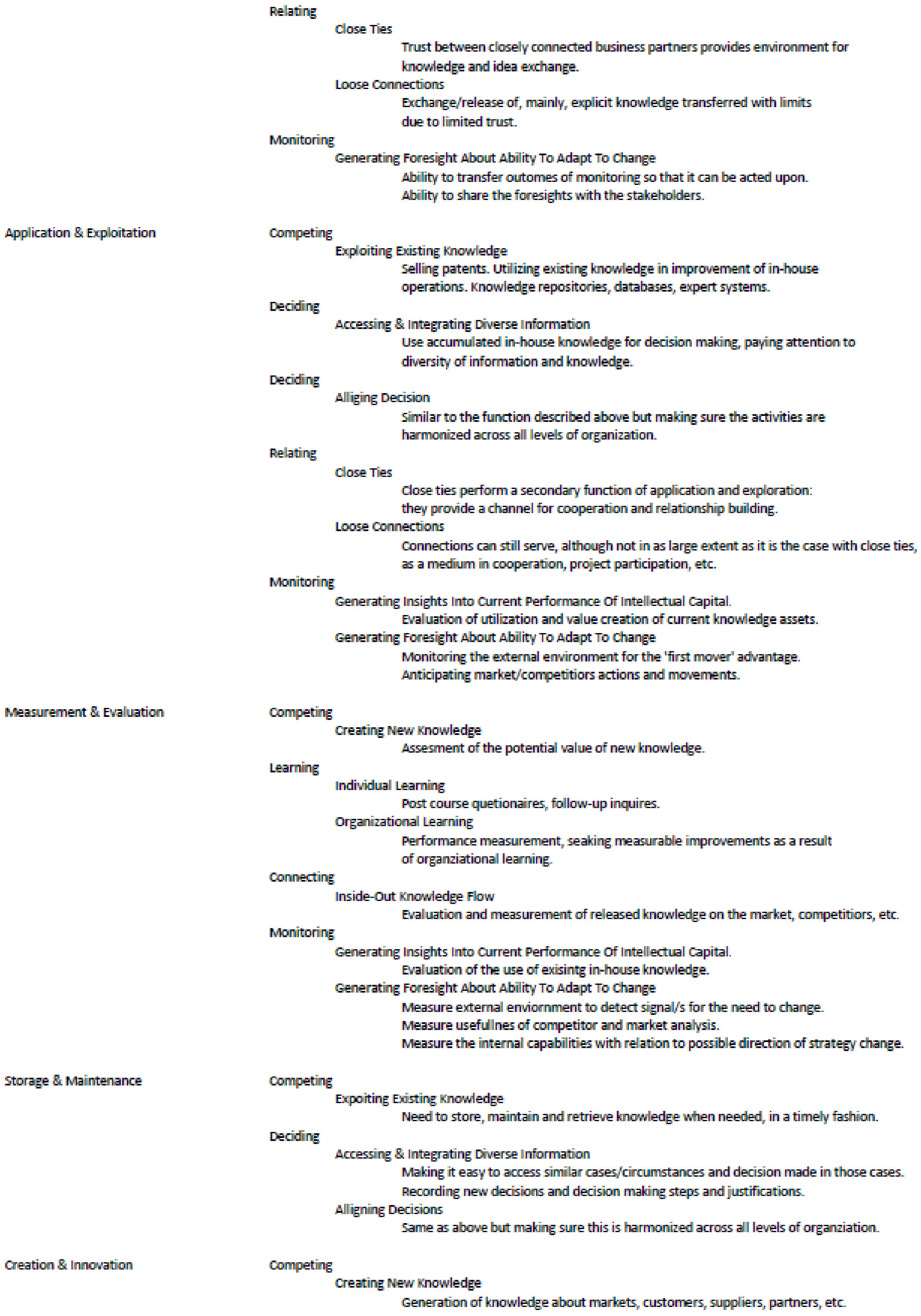
Connecting

Outside-In Knowledge Flow

Knoweldge needed for environmental scanning needs to be transferred to the organization and, perhaps, propegated further within the company.

Inside-Out Knowledge Flow

Release of information expected to benefit the organziation upon release.



	Relating	Close Ties	Trust between closely connected business partners provides environment for knowledge and idea exchange.
		Loose Connections	Exchange/release of, mainly, explicit knowledge transferred with limits due to limited trust.
	Monitoring	Generating Foresight About Ability To Adapt To Change	Ability to transfer outomes of monitoring so that it can be acted upon. Ability to share the foresights with the stakeholders.
Application & Exploitation	Competing	Exploiting Existing Knowledge	Selling patents. Utilizing existing knowledge in improvement of in-house operations. Knowledge repositories, databases, expert systems.
	Deciding	Accessing & Integrating Diverse Information	Use accumulated in-house knowledge for decision making, paying attention to diversity of information and knowledge.
	Deciding	Alliging Decision	Similar to the function described above but making sure the activities are harmonized across all levels of organization.
	Relating	Close Ties	Close ties perform a secondary function of application and exploration: they provide a channel for cooperation and relationship building.
		Loose Connections	Connections can still serve, although not in as large extent as it is the case with close ties, as a medium in cooperation, project participation, etc.
	Monitoring	Generating Insights Into Current Performance Of Intellectual Capital.	Evaluation of utilization and value creation of current knowledge assets.
		Generating Foresight About Ability To Adapt To Change	Monitoring the external environment for the 'first mover' advantage. Anticipating market/competitiors actions and movements.
Measurement & Evaluation	Competing	Creating New Knowledge	Assesment of the potential value of new knowledge.
	Learning	Individual Learning	Post course quetionaires, follow-up inquires.
		Organizational Learning	Performance measurement, seaking measurable improvements as a result of organziational learning.
	Connecting	Inside-Out Knowledge Flow	Evaluation and measurement of released knowledge on the market, competitiors, etc.
	Monitoring	Generating Insights Into Current Performance Of Intellectual Capital.	Evaluation of the use of exisintg in-house knowledge.
		Generating Foresight About Ability To Adapt To Change	Measure external enviornment to detect signal/s for the need to change. Measure usefullnes of competitor and market analysis. Measure the internal capabilities with relation to possible direction of strategy change.
Storage & Maintenance	Competing	Expoiting Existing Knowledge	Need to store, maintain and retrieve knowledge when needed, in a timely fashion.
	Deciding	Accessing & Integrating Diverse Information	Making it easy to access similar cases/circumstances and decision made in those cases. Recording new decisions and decision making steps and justifications.
		Alligning Decisions	Same as above but making sure this is harmonized across all levels of organziation.
Creation & Innovation	Competing	Creating New Knowledge	Generation of knowledge about markets, customers, suppliers, partners, etc.

Fig. 4.6.2.2: The mapping between the KM-process-based model of Burnett et al. (2004; 2013) and the six competence model of McKenzie and van Winkelen (2004)

Section 3.2.3 listed other KM frameworks that were considered in this research; they are not repeated here.

4.7 Ethics

As per the specifications of a research project by Saunders et al. (2009, pg. 137), which called for an investigation of ethical aspects, the ethical issues associated with this research are considered next.

Walliman (2011, pg. 240) identifies the following ethical issues that should be considered when conducting research:

- Honesty in work (properly addressing intellectual ownership, plagiarism, citation, acknowledgments, data interpretations and assumptions based on epistemology).
- Situations that raise ethical questions (these include research aims – are there any consequences in relation to them?; ethics in relation to other people or organizations; potential harm and gain that results from the research conducted).

The aim of this research is to add to the body of knowledge that attempts to illustrate how DM can be used as an analytical tool for evaluating the impact of KM processes on OR. Considering the aims of this research, it can be concluded that that it has little or no direct ethical consequences. There are, however, other ethical factors to address.

With regard to the research participants, what Creswell (2009, pg. 91) refers to as the protection of the anonymity of individuals was ensured. The names of the participants or their organizations and/or any other information allowing for the identification of a respondent (such as an IP address) were substituted by a participant sequence number (consisting of consecutive numbers 1 through to 46). All references were then made to the assigned consecutive number, thereby concealing the real identity of the respondent. (The IP addresses of respondents were not stored in the database that held the replies to the questionnaire. Other than IP addresses, the replies contained no information regarding the respondents.)

To encourage responses, the questionnaire never asked specific performance/financial questions. Instead, the questions dealing with the performance or financial standing of an organization were asked in relative terms, relating to some period in the recent past.

While it is difficult to foresee the indirect impact of the research, one area can foreseeably affect people and their roles. It is possible that, as the result of this research, DM tools could gain wider use in the evaluation of the impact of KM processes, allowing organizations to focus on the 'most important' KM processes at the expense of other KM processes. This could possibly lead to a positive impact on one group of people (those involved in 'high-valued' KM processes) at the expense of another group (those engaged in the 'low-valued' KM processes). However, such an outcome can be compared to the market forces that dictate 'premiums' for certain roles over others.

With regards to the data generated by the research, it was never shared with anyone and is stored on a device that is protected by commercial software.

The introductory letter sent along with the questionnaire to the recipients informed them about the nature of the study they were asked to participate in and assured the anonymity of their replies.

Finally, in order to encourage completion of the questionnaire, the introductory letter to the questionnaire participants offered a comparison of the respondent's organization with the responses from all of the other respondents and an electronic copy of this thesis. While the low return rate in response to the questionnaire does not warrant presentation of extensive and complete analysis, some feedback and a copy of this thesis will be provided to organizations that request them. It is the hope of the author of this work that the electronic version of this thesis will introduce new ideas concerning OR, KM and DM to the organizations that have participated in this research, leaving them 'better off'.

4.8 Methods

Saunders et al. (2009, pg. 3) define a method as a technique or procedure used to obtain and analyze data. This section discusses research methods in the context of this work; it is a continuation of the discussion of methods that began in Section 4.2.2.

Some commonly utilized research methods, as stated by Rajasekar et al. (2006, pg. 2), include theoretical procedures, experimental studies, numerical schemes and statistical approaches. The more common and perhaps basic classification of research methods, which has been used by many writers, classifies the research methods into three basic types: qualitative, quantitative and mixed (Creswell, 2003; Saunders et al., 2009; McCusker & Gunaydin, 2015). Based on the work of McCusker and Gunyadin (2015, pgs. 537-540) and Saunders et al. (2009, pgs. 151-155), the characteristics, advantages and disadvantages of each of the three research methods are presented below.

Qualitative research tends to answer the questions 'how', 'what' and 'why'. It is used mainly as a data collection technique (through conducting interviews) or a procedure for analysis (through categorizing data). The quality of the research matters greatly, as the research in fact becomes the researcher's tool, including the philosophical perspective of the researcher. This allows many factors to be investigated and also provides a context for the responses provided by the data collection instrument used in the research, which potentially leads to a deeper understating of the responses. Considerable time is required for data collection, and there are possible ethical issues related to the information collected. McKusker and Gunyadin (2015, pg. 539) point out that qualitative research is often used prior to quantitative research; usually, it is used in the initial stages or while validating the idea behind a research project.

Quantitative research often answers the questions 'how many' or 'how much' when investigating a phenomenon. It is also used as a synonym for data collection (the use of questionnaires) or as a data analysis procedure (statistics, graphs and/or charts) that generates or is based on numbers. The objective of this type of the research is typically to count features/events and classify them in order to explain the subject observed. In quantitative research, the researcher knows in advance what he or she is looking for, and the study is well designed prior to the application of the quantitative methods. The researcher tends to be objective about the researched topic, and the focus is on the generalization of findings.

Mixed methods research utilizes the strengths of both qualitative and quantitative data collection and analysis techniques in one research project, at the same time, but with a clear separation between them. 'Mixed methods can provide significant pragmatic advantages when exploring complex research questions' (McKusker & Gunyadin, 2015, pg. 541).

A research approach should be selected based on the requirements of the author. 'The reasons for choosing particular data collection and analysis methods are always determined by the nature of what you want to find out, the particular characteristics of research problem and the specific sources of information' (Walliman, 2011, p. 173).

Additional methods have been employed for the analysis of the data for the purpose of DM; these methods are discussed in Section 5.3. The tools for the analysis of data described in Section 5.3 were chosen based on their availability and the author's familiarity with them. The software products used included Microsoft Excel (V. 2010), Easy Fit 5.6 Professional and MaxStat Pro 3.6, with the majority of work being conducted in Excel. The application of these tools is described in Section 5.3.

Having decided on the type of research, the next choice to make is that of research design. Due to the non-experimental nature of the research, the factorial family of designs and multiple-group designs were excluded from consideration. From the one-group design family, which is the family that includes the most common pretest-posttest design, the interrupted time series design and the correlation design, the correlation design appears to be the best choice of research design. The cross-sectional design was selected as the most appropriate for the task at hand, as the task involves determining the relationships between OR and KM processes. Because of the lack of prior research in the area, the consequent uncertainty regarding the results obtained and time constraints, this first study was limited to merely making an observation on all variables at one point in time. The notation used to represent this research design, from the now-classic work of Spector (1981, pg. 27), is, therefore, 'O', where 'O' represents all observations on all variables. Additional discussion about cross-sectional design takes place Section 4.8.2.

This work follows the quantitative research method, as a questionnaire is used to collect data that is later analyzed using DM tools. This selection of method (quantitative for both data collection and analysis) is in-line with the philosophical perspective that guides this research (post-positivism) per McKusker and Gunyadin (2015, pg. 540), who identify quantitative research with positivism based on the objective approach being derived from the quantitative research. The selection of the quantitative method was also driven by the need to count events (KM activities/group of activities) or features in order to explain corresponding numerical OR values (which also needed to be captured and be measured, in a way that is similar to the six competence area approach, across all respondents).

4.8.1 Data: Primary vs. Secondary Sources

The selection of the data sources is one of the most important aspects of research. The two choices for the data sources are primary and secondary sources, with each having its own characteristics, limitations and possibility of being affected by potential errors (Rabinski, 2003, pg.1).

In the most elementary classification, the primary data sources are sources that involve the collection of data directly from an organ of interest (an organization or individual, for example), while the secondary data sources refer to data obtained from publically available sources.

Surveys, interviews, tests, experiments, accounts and observations are all examples of the collection of data from primary data sources, while written material (publications, letters, reports and books, for example) as well as non-written material (such as works of art, historical artifacts and recordings) constitute secondary sources. Clearly, the examples presented are not intended to form an exclusive list of the elements that constitute, for example, written materials. An important generalization to make is that any source that was used indirectly by a researcher can be labeled a secondary data source.

Some main advantages of using primary data include the following:

- The ability to capture intangible information and collect more complex data;

- Custom data collection design allows the data to be better aligned with the needs of the research;
- Collection of publically unavailable data – primary sources are often the only source of data, especially for small and mid-size enterprises; and
- Learning opportunities are associated with the development of the data collection device.

Some main disadvantages of using primary data are as follows:

- The possibility of unrepresentative samples and other related issues;
- The risks associated with the development of researcher's own measuring device; and
- The monetary and time costs of carrying out such data collection.

While secondary data, mainly concerning financial and operational matters, is widely available for public US companies, data about mid-sized, private US companies is, to a large extent, kept secret (and since, as mentioned in Chapter 1 and Section 4.8.6, this research focuses on mid-sized companies, this makes the use of secondary data a poor choice for this research). Therefore, the data for this research could only be obtained from primary sources. Moreover, from the author's nearly 18 years of personal experience dealing with the management of the mid-sized US companies in the capacity of software/technology consultant, one observation that should be noted is that the questions presented to the management of such companies should be somewhat general to ensure that they do not give rise to privacy and strategic concerns. Similarly, the questions asked on behalf of this work needed to be governed by the same principle of preserving privacy, and there was a need for an introductory letter that assured the confidentiality of data collected. Not adhering to such practices raised the risk of low reply rate and biased answers. More information on these aspects of this research can be found in Section 4.8.8.

Due to the importance of primary data sources to this research, issues associated with the primary data source are discussed in greater detail next. Some of the errors discussed next can also be present when primary data is aggregated and presented as secondary data; however, the emphasis of the discussion is on primary data sources. The work described below is based on that of Rabinski (2003, pg. 48).

According to Rabinski, 'When primary data is generated by either observation or questioning, the resulting data contains whatever bias and error arose in the process of data gathering.' Also according to Rabinski, there are two terms that are used when discussing the full extent of the issues associated with data handling: 'sampling' and 'non-sampling' errors.

A sampling error is an error directly related to the selection of the population sample. It occurs when the chosen sample does not accurately reflect the total population under investigation. The assumption of this research is that, for practical reasons, the total population will be used; thus, this type of error is not applicable to this research.

A non-sampling error arises during the measurement process, after the sample of the population has been determined (Rabinski, 2003, pg. 48). There are five general non-sampling errors that can occur at this phase:

- Frame error – This occurs when the list that the analyst generates to represent the population omits certain individuals whose opinions, attitudes, or other characteristics will not otherwise be represented. (For the purpose of this research, it is unclear if all respondents of lower than CEO rank shared the CEOs' opinions.)
- Measurement error (aka response error) – This arises when the individual who responded to the questions gives information that is not true. It also occurs when the analyst misrepresents observable facts. (This study made no attempt to use any of the techniques that attempt to detect conflicting responses.)
- Sequence bias – This occurs when the order of questions on an administered measuring instrument suggests or induces an idea or opinion in the mind of the respondent as a direct consequence of the manner in which the question was phrased. (The order of questions in the questionnaire was based on the order of competence areas presented in McKenzie and van Winkelen model and appears not to induce any opinions.)
- Interviewer bias – This occurs because of the presence or influence of an interviewer in a person-to-person interview. (This is not applicable to this research due to its quantitative nature.)

- Non-response bias – This occurs when the individuals in the sample do not respond to some or all of the questions or fail to participate in the study. (There were a few responses that were abandoned, as their completion level was less than 10% of the questions, making the replies unusable for the purpose of this research.)

The primary data collection instrument selected for this research was a questionnaire; justification for such a choice was provided in Section 4.8.3 when expanding on the discussion of the measuring instrument.

4.8.2 Cross-sectional vs. Longitudinal

From the detailed procedures of data collection and analysis, or what Creswell (2009, pg. 145) refers to as 'methods', it is necessary to decide on the source of data, which primarily involves making a choice between primary and secondary data and deciding on the specific data analysis tool to be used. A discussion of the data sources used in this research takes place in Section 4.8.1.

The time horizon is another aspect that must be taken into consideration when planning research. As stated by Saunders et al. (2009, pg. 155), the time horizon of the research, independent of the research methodology and methods, depends on whether the research is done at one point in time (which is referred to as cross-sectional) or is in a form of a diary (which is referred to as longitudinal). Most doctoral research projects, according to Saunders et al., due to time constraints, are of the cross-sectional type.

As noted by Spector (1981, pg. 33), the cross-sectional approach is used in determining if two or more variables are related; the establishment of such a relationship is often, in itself, the research question.

Levin (2006, pg. 24) provides further insights into the cross-sectional approach by making the following statements:

- The purpose of the study is descriptive, often being a survey. There is typically no hypothesis as such, with the primarily goal being to describe a population with respect to some outcome.

- The purpose of the study is to determine the prevalence of the outcome of interest for the population or part of it, at some point in time.

The advantages and disadvantages of cross-sectional studies, based on the work of Levin (2006, pg. 27), are summarized in Table 4.6:

Advantages:	Disadvantages:
Does not consume much time to conduct. Relatively inexpensive.	Difficult to make casual inference.
Because a sample is usually taken from the entire population it can estimate prevalence of outcome.	Provides data in the form of a snapshot: different results will perhaps be obtained if another time-frame is chosen.
Numerous outcomes can be assessed.	Prevalence-incidence bias is present.
No follow-up study is required.	

Table 4.8.2.1: Advantages and disadvantages of cross-sectional studies

The longitudinal design's main strength is 'the capacity that is has to study change and development' (Saunders et al., 2009, pg. 155). In a sense, observations over a long time frame provide great ability to control the observed variables of a study, provided such observation does not affect the research itself.

Because of the emphasis of this research on determining the relationships between variables and the need to investigate numerous possible outcomes of such relationships, the cross-sectional approach is used in this thesis. The use of the longitudinal approach with DM as the analysis tool is probably not justifiable due to the complexities involved. (Chapter 5 implicitly assumed the use of single numbers instead of, for example, the number vectors that could represent the longitudinal measurement.) Moreover, in practice, DM models are re-built many times over a period of weeks/months so that the latest data can be incorporated into the model; otherwise, the models would become 'stale', making the measurements 'somewhat' longitudinal anyway.

4.8.3 Questionnaire as Data Collection Instrument

‘Questionnaire - General term including all data collection techniques in which each person is asked to respond to the same set of questions in a predetermined order’ (Saunders et al., 2009, pg. 599).

Measured Items

The questionnaire research strategy tends to be used primarily in exploratory and descriptive research that attempts to centrally answer the questions of who, what, where, how much and how many questions (Saunders et al., 2009, pg. 144). While this work does not seek to answer any of these questions directly, the type of the research is, indeed, somewhat exploratory (however, this research is applied, rather being basic/exploratory), and it uses its 84-question questionnaire as a data collection instrument, which is later analyzed using the DM. (More details on the choice of measuring instrument are provided in section that follows and called: Questionnaire used by this work.)

Range of Scales

Rattray and Jones (2007, pg. 235) provide a list of some of the most common ranges of scale used when developing questionnaires that includes frequency scales, the Thurston scale, Guttman scaling, Rasch scaling and the Likert scale, with the Likert-style scale (with a varied number of points on the scale) being the most commonly used. Dennis (School of Public Health & Community Medicine) notes that the ideal number of options on the Likert-scale is either five or seven. Converse and Presser (1986, 37) state the following: ‘Two of the most commonly used intensity indicators are “strongly agree, agree, disagree or strongly disagree” items.’

Typically, according to Rattray and Jones (2007, pg. 236), the Likert-scale assumes that the intensity and/or strength of experience is linear (that is, it can be expressed on the strongly disagree to strongly agree scale) and that attitudes can be measured, with the most commonly used five, seven and nine element scales including a neutral point.

Somewhat of an open issue for discussion is the inclusion of the neutral option as an answer. Rattray and Jones (2007, pg. 236) point out that the lack of the neutral option in a questionnaire may aggravate respondents if they find the answers they are forced into giving not applicable to their situations.

The author realizes that, given the target organizations and varied industries involved, there might be many situations for which a given questionnaire question will not be applicable. To resolve this problem, without affecting the results obtained from the questionnaire, the 'not applicable' ('N/A') option was added to each question. (A description of how such responses are handled without affecting the other questions within the questionnaire or other questionnaires is provided in Chapter 5.)

Design Considerations

While this research is not focused on drawing conclusions from the data received, it does use the questionnaire as the data collection method and, as such, a few important details with regard to questionnaire design require attention. Because the issues related to survey design are, in large part, applicable to the questionnaire design used in this research, such issues are presented in Table 4.8.3.1 with brief annotation, in parentheses, about their applicability to the instrument used in this research.

Fowler (2014, pg. 75), Converse and Presser (1986, pgs. 9 -47) and Loughborough University's (2011) handout identify out several items to consider concerning the construction of a survey. These items, which have not been discussed thus far, are presented in Table 4.8.3.1, below:

Aspect of survey (also applicable to questionnaire) design to consider:	**Justification:**
Simple language (Considered in questionnaire design.)	To ensure that the survey recipient, who is often less educated than the survey author, clearly understands the question.
Short questions (Considered in questionnaire design.)	To allow the reader to remain concentrated and remember what s/he is asked about.
Avoid confusion and do not: use ubiquitous questions, double negatives	All of these aspects make the survey difficult to understand or provide room

in a question, implicit negatives, overlong lists or asking the question before introducing the topic. (Considered in questionnaire design.)	for ambiguity and loss of the respondent's interest and focus.
Common concepts/shared definitions (Considered in questionnaire design.)	To provide an explicit definition or common frame of reference so that the respondent's common concepts match those of the survey author.
Recalling the past in a question (Considered in questionnaire design. Narrow, memorable in history, period used for comparison.)	Do not use unless necessary as memory questions tend to be difficult. If used, narrow the period of recall.
Hypothetical questions (Not applicable.)	Avoid these as they tend to also produce hypothetical answers.
Use of stories (Not applicable.)	Due to their length, their number should be limited. Care should be taken that the reader does not become bored with the story based on which the question/s is/are derived.
Specific questions are better than generic ones. (Considered in questionnaire design.)	To ensure the topics asked about have the same meaning to other people. Provide better recollection.
Open/closed questions (Considered in questionnaire design.)	Closed questions (allowing the selection of alternative answers) provide the same frame of reference to all respondents.
Forced-choice questions/Not agree-disagree statements (Not applicable.)	Forced-choice items are more apt to encourage a considered response than agree-disagree statements.
Order of questions and wording (Considered in questionnaire design.)	Need to be considered, as they may bias response. Loughborough University's questionnaire-design handout proposes moving from general to specific questions, from factual to

	abstract questions, from closed to open questions and leaving the demographic and personal questions for last.

Table 4.8.3.1: Survey construction - items of consideration

Questionnaire Used By This Work

The questionnaire used in this research, presented in Appendix I, has been constructed to measure several aspects of KM and the impact of KM on OR. In the most simplistic categorization, the questionnaire measures two things: independent and dependent variables and the impact of the independent variable (KM) on the dependent one (OR).

In more detailed terms, the questionnaire attempts to measure the following:

- The extent to which a given organization utilizes KM processes, the independent variables. (Those can be later grouped into logical categories, such as competence areas). These are the questionnaire's questions 1 through 52.
- The performance of an organization (questions 53 through 68) in order to achieve the following goals:
  - Ensure the organization operates in a non-declining industry, so that the responses are not affected solely by a negative business environment.
  - Correlate the performance of an organization to its OR (for the sake of the argument stated in Chapter 3.)
- Organizational resilience, the dependent variable (questions 69 through 84), to assess how resilient an organization truly is.

As noted in Chapter 5, the individual answers to the questions contained in the questionnaire may need to be grouped into logical groupings based on the DM algorithm used. The six competence areas provide one of the possible groupings; others, including ones generated as an outcome of the segmentation algorithm, are also possible.

As a result, a questionnaire consisting of 84 questions divided into eight sections was created to serve as the data collection instrument. The on-line questionnaire was preceded by a half-page introduction to the research and the researcher. The questions within the questionnaire were combined into sections according to the McKenzie and van Winkelen model: 1) competing, 2) deciding, 3) learning, 4) connecting, 5) relating, and 6) monitoring, with each section clearly described in the online questionnaire. The questions within each section of the questionnaire were derived from the corresponding section of the McKenzie – van Winkelen model (2004). The competing section was further split into two implicit sub-sections: one relating to the new knowledge and the other based on the exploitation of existing knowledge. In addition to the sections based on the McKenzie and van Winkelen model, there were two additional sections: one section related to operational performance and one to OR. The questions within the OP and OR sections were based on the literature review, with a particular focus on the authors whose definition of OR is used in this research: Hamel and Valinkangas (2003), with additional insights from Robb (2000) and Mallak (1998), among others. Finally, within the OR section, there were six questions (#77-#82) that attempted to determine whether an organization operates in a declining industry, in which case the replies would have to either considered separately or be excluded from consideration.

The preference for the use of the Likert scale in the measurement instrument is due to the fact that the scale allows for each question to measure the intensity, or the measure of strength, of the answer provided by the responding party. Such a measure of intensity can provide an additional insight into the collected data (Converse & Presser, 1986, pg.37), which could be highly valuable to this research. The intensity indicators used by respondents are presented below, and they consist of the following, which, according the Converse and Presser quote above, are mostly commonly used in such questionnaires: 'strongly disagree', 'disagree', 'agree' and 'strongly agree'. In addition, neutral answers 'neither agree nor disagree' and the 'not applicable' options were provided as well. Fig. 4.5 is an illustration of the scale used in the instrument. (The five-element Likert-scale, with the 'N/A option,' was ultimately selected for this research, as using a smaller number of elements would not capture what felt like an existing distinction between too narrow choices. The nine-element scale was not used, as it would look 'too busy' given the number of questions asked in the questionnaire.)

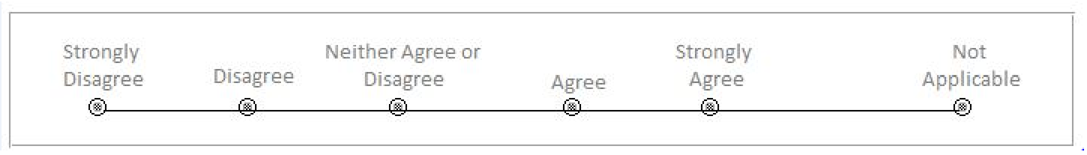


Fig. 4.8.3.2: Likert-scale used in the questionnaire utilized by this research

The design of the questionnaire was influenced by the work of Fowler (2014), Rattray and Jones (2007), Converse and Presser (1986), Carmines and Zeller (1979), Spector (1981) and Frary (2003). As the first step towards the creation of the questionnaire, it was presented to the supervisor of this research, Professor Burnett; it was then mailed, via mail with a pre-paid return envelope and an introductory letter, in late September 2013, to five mid-sized organizations in the mid-western part of US as a pilot study (discussed further in Section 4.8.6), followed by a reminder to complete the pilot questionnaire that was sent four weeks later.

With the exception of one returned questionnaire that asked for a clarification of a single question (#65), the remaining questionnaires that were received contained comments that the questions were concise and clear and had no suggestions for improvements. (Question #65 was later modified to reflect the suggestion received from the pilot study.) Moreover, all replies stated that the size and complexity of the questionnaire were manageable and that the completion time of 30 minutes was accurate. (Note that, during the data processing phase. the responses to question # 76 were omitted in analysis, as this was the only question to which the 'strongly agree' reply did not necessarily correspond with the most desired option for an organization. That is, the 'Strongly Disagree' option was the answer option to be selected by the resilient organizations as the answer to this question, in contrast to all other questions to which the answer 'Strongly Agree', was the answer expected from the resilient organization. This means that it was not clear how to assign a numerical value to the response to that question, as 'strongly disagree' would decrease the sum of points collected to all other questions.)

The questionnaire and the introductory letters used in this research are presented in Appendices I and X.

Side note: The mailing, in addition to the letter, contained a business card and a link to the www.surveymonkey.com site where the questionnaire could be completed. (The choice of mailing, rather than emailing, the invitation to complete the research questionnaire was based on two main reasons: When speaking with the executives of seven mid-size organizations about their email habits, the impression was that it would be very unusual for an executive to open an email from a stranger and even less likely that they would click on the link contained within such an email; second, the costs of obtaining executives' email addresses would make the purchase of the contact lead 6.25 times more expensive.)

4.8.4 Reliability & Validity

Some of the most important aspects of successful research are reliability and validity. The concept of reliability is discussed next, followed by a discussion of validity.

The stages presented in Fig. 4.8.4.1 must occur in order for a question to be reliable and valid.

Carmines and Zeller (1979, pg. 12) define reliability as a tendency toward consistency. 'The more consistent the results given by repeated measurements, the higher the reliability of the measuring procedure.' Reliability, therefore, focuses on the extent to which the results are consistent through repeated measurements of the same object.

Reliability and validity are greatly affected by random (occurring by chance) and non-random errors (such as those that result from a systematic biasing effect on the measurement instrument)

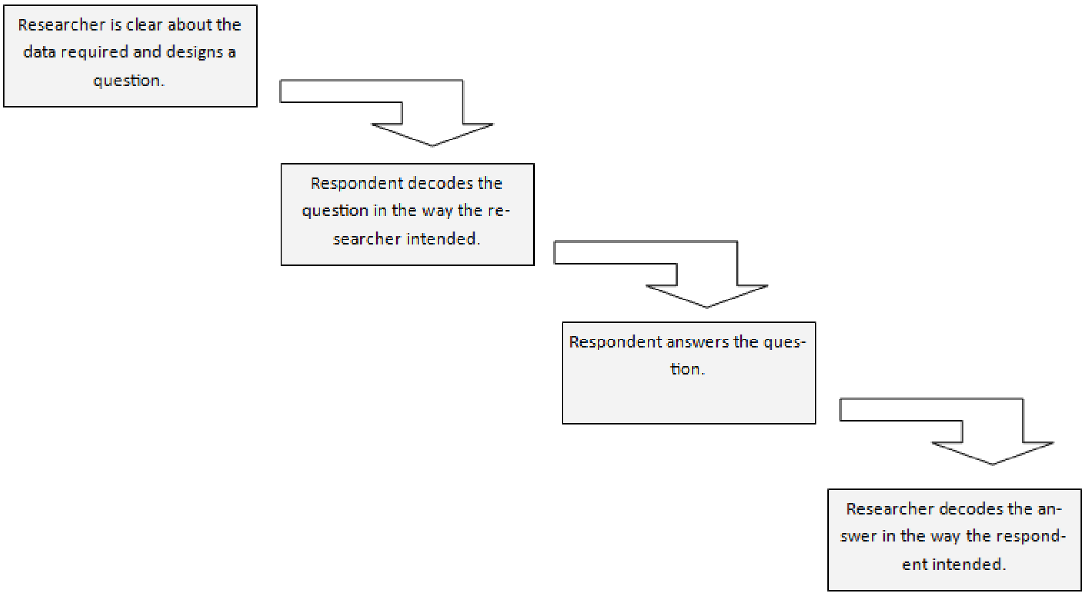

Researcher is clear about the data required and designs a question.
Respondent decodes the question in the way the researcher intended.
Respondent answers the question.
Researcher decodes the answer in the way the respondent intended.

Fig. 4.8.4.1: Survey reliability and validity. [Derived from Saunders et al. (2009, pg. 372).]

'Just as reliability is inversely related to the amount of random error, so validity depends on the extent of nonrandom error present in the measurement process' (Carmines & Zeller, 1979, pg. 15).

Since the focus of this research is not on the numerical outcome of the questionnaire but rather on the methodological establishment of DM-based methods, the discussion of the estimation of error and/or reliability of the empirical measurement is brief. This discussion, which takes place next, mainly addresses reliability and Cronbach's alpha, which was used in this research to measure instrument reliability.

Carmines and Zeller (1979, pgs. 37-52) discuss four methods of estimating the reliability of empirical measurement. The methods discussed include the following:

- Retest method – the same test is applied to the same object at some later time and correlations are then obtained between the scores of the same test;

- Alternative-form method – similar to the retest method, but rather than giving the same test a second time, an alternative form of the same test is administered;
- Split-halves method – while the prior two methods require the administration of two tests, this method divides the total set of items into two halves and the scores from each half are correlated to estimate the reliability. Unfortunately, the correlation between the halves will differ somewhat based on how the total number of items is divided; and
- Internal consistency method – provides a single administration of the test that results in a unique estimate of its reliability, calculated via mathematical formula. Cronbach's alpha, according to Carmines and Zeller (1979) and other writers (Field (2006), Gliem & Gliem (2003), Icabucci & Duhachek (2003)) is the preferred method for assessing the reliability of a measurement.

Based on the discussion above, this research uses Cronbach's alpha to calculate the reliability of each segment of the questionnaire (each competence area). The calculation was accomplished via a downloadable resource pack from www.real-statistics.com that extends Microsoft Excel by adding the Cronbach's alpha function to it.

Cronbach's alpha, a widely used indicator of a measurement's reliability, was assessed (Isik et al., 2013, pg. 19), and exploratory factor analysis was used to assess dimensionality. Moreover, an assessment of the construct's validity of dependent and independent variables was performed, employing conversant and discriminant validity. As described by Isik et al. (2013, pg. 19,) convergent validity was assessed according to average variance extracted (AVE) and communality. Discriminant validity was assessed by comparing the square root of AVE with each construct and inspecting to see if the square root was of greater value.

Per Carmines and Zeller (1979, pg. 51), since Cronbach's alpha needs to be computed for any multiple-item scale, for the purpose of this research, Cronbach's alpha was computed for competence area. The results of the computation of Cronbach's alpha are presented in Table 4.8.4.1.

Field (2006, pg. 1) states that a value of 0.7 – 0.8 is an acceptable value for Cronbach's alpha to indicate reliable scales.

Questionnaire's section short name:	Cronbach's alpha:
Create (CR)	0.67
Exploit (EX)	0.60
Decide (DE)	0.76
Learn (LE)	0.75
Connect (CO)	0.79
Link (LI)	0.80
Performance (PE)	0.83
Organizational Resilience (OR)	0.83
All fields of questionnaire at once	0.95

Table 4.8.4.1: Cronbach's alpha scores for the questionnaire and various sections of it used in this research

Carmines and Zeller (1979, pg.17) quote Cronbach (1971, pg. 447): 'One validates, not a test, but an interpretation of data arising from a specific procedure' when discussing validity. The reason for the need for such a statement is the fact that it is not the validation of the instrument that is needed, as the instrument can still be valid but measure inappropriate phenomenon, but the validation of the instrument in relation to what it is supposed to measure.

The validity of the data collection instrument (in the case of this research, the questionnaire) can typically be assessed by the following four methods (Saunders et al., 2009, pgs. 372-373):

- Internal validity – the ability of the measurement instrument to measure what it was designed to measure;
- Content validity – determining whether, and the extent to which, the instrument provides adequate coverage of investigative questions;
- Criterion-related validity – also known as 'predictive validity,' is concerned with the ability of the measure (the questionnaire's questions) to make accurate predictions; and

- Construct validity – refers to the extent to which the questionnaire questions actually measure the presence of the constructs intended to be measured.

Finally, in order for a measure to be concept-validated, there must exist a theoretical network supporting the concept (Carmines & Zeller, 1979, pg. 23), and special care must be taken when the construct-validity is negative as that can indicate one of the following issues: the indicator does not measure what it purports to measure; an incorrect theoretical framework was used to generate the empirical prediction; a faulty or incorrect method or procedure was used to test the theoretically-based hypothesis; or, there is a lack of construct validity or another variable(s) included in the analysis is unreliable.

Because of the nature of this research, which focuses on the establishment of a methodology for measuring on the impact of KM on OR using DM, the use of the questionnaire is a part of the methodology, as the data from the questionnaire is processed for illustrative purpose. With this in mind, there was no need to test the validity or reliability of the measuring instrument. Despite that, however, the reliability of the questionnaire has been measured and the outcomes are reported in this section.

4.8.5 Sampling

One of the first steps in survey design, according to Kalton (1983, pg. 6), is arriving at a definition of the population to be studied.

Anderson et al. (2003, pg. 14) define a population as follows: 'A population is the set of all elements of interest in a particular study.'

This research reached out to the entire population, as described below and in Section 4.8.6, so the aspects related to the sampling techniques and topics related to sampling (such as representativeness and quality, sample size, estimation error and others) are omitted from the discussion.

The sample size was the entire population used in this research.

To conduct this study, 3,413 companies were sent, via regular mail, an invitation to complete an on-line questionnaire; this constitutes not a sample but the entire population, given the definition of the mid-sized company and the geographic area considered in this research. One can expect Dun & Bradstreet, the source from which the list of the companies was obtained, to provide information on all mid-sized companies in the mid-west region.

Another point to consider when using a questionnaire as a data collection instrument, in addition to the issues previously discussed in this section, is the issue of non-response – in the sense that non-respondents will differ from respondents, thereby introducing bias to the measurement. Because of the focus of this research on methods instead of numerical outcomes, this issue did not have to be addressed.

Isik et al. (2013, pg. 20) also discussed the issue of 'expansion of the sample size,' which at one point was a very important aspect for this research, as this study considered expansion of the replies by increasing the number of questionnaire replies via the methods used by Isik et al.: They increased sample size by generating 500 random samples from the survey replies. The bootstrapping procedure available in the SmartPLS Software was used by Isik et al. (2013, pg. 20) for such expansion of sample size.

4.8.6 Sampling Strategy and the Selection of Firms

There are several reasons behind the selection of the mid-sized companies and mid-western part of the USA investigated in this research. From the personal perspective of the author of this work, who worked as an independent BI professional and has over twenty years of experience serving mid-sized companies in the mid-western part of the US, this area provides very familiar ground for an investigation and the business-based reasons.

As stated by Fink and Ploder (2007, pg. 705), traditionally, KM focused on the domains of larger organizations, and the aspects of culture, networking, organizational structure and technology infrastructure tend to be applied to large multi-national organizations and are given little relevance to the small and mid-size companies (SMB). However, Fink and Ploder state (2007, pg. 705) that the success of SMBs depends on how well such organizations manage the knowledge of their knowledge workers. The challenges facing small and mid-sized companies have been also discussed by Kipley et al. (2008, pg. 18). According to Kipley et al., SMBs still view KM as a vehicle only for efficiency improvement, rather than seeing it as a vehicle for the improvement of corporate functionality. Because of the importance of the KM initiatives for mid-sized companies and the need for mid-sized companies to look past operational efficiencies in order to flourish in today's complex business environment, this study has been undertaken. It is expected that, as result of this study, DM-based methods for establishing the relationships between KM initiatives and organizational performance/resilience will be practically developed. Such methods could be later utilized to validate the importance of KM for OR within SMB organizations.

According to Fink and Ploder (2007, pg. 706) SMBs 'do not have much money to spend on knowledge management initiatives, so knowledge must be leveraged so that goals can be achieved in an effective and efficient manner.' Being able to identify, through this research, the most influential KM processes that add the most value to mid-sized companies can lead to better utilization of the financial resources of mid-sized US firms.

Mid-sized companies are typically underestimated in terms of their impact on the US economy. According to Deloitte's 2012 mid-market perspectives report, the mid-market companies employ, as a group, more people than the entire S&P 500 and have total revenues equivalent to 40 percent of the US GDP. With such a significant portion of the US GDP generated by these mid-sized companies, it is hypothesized that even a small improvement in organizational performance achieved through KM will have significant effect on the revenues and portion of US GDP contributed by mid-sized companies.

Similarly, the Midmarket Institute states that '[m]idsize companies account for just 3.2% of all companies in the U.S. and yet provide 34% of all jobs, 31% of all US revenue.

The mid-west region includes the following states: Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota and Wisconsin.

Definition used in the research and justification for such a choice.

The CNBC Corporation defines mid-sized companies as those with annual revenues of between $10 million and $1 billion.

USA Today uses revenues between $10 million and $1 billion as a range for defining mid-sized companies, also stating that there are about 200,000 such firms in the US as of 2012.

CNN, in their 2012 'Survey of best mid-sized companies to work for,' classified mid-sized companies as organizations employing between 2,500 and 10,000 full-time employees.

There is slight confusion regarding what type of organizations constitute mid-size companies in the US, and there are a number of terms used. It appears that every person or organization has their own definition of such a firm. The U.S. Small Business Administration, as opposed to its European counterpart, the EU's European Commission, does not provide precise headcount sizes or revenues regarding what constitutes a mid-sized US company. Similarly, the U.S. Census Bureau only goes so far as to provide various bands (firms with 750 to 999 employees as band 'C,' etc.) for its own reporting purposes and does not specify what constitutes a large or small firm (smbresearch.net/sizing-up-smb/). If there is any consensus at all, it might be perhaps best represented by the work posted on the web by Gartner, which defines a mid-size organization by stating that '[SMBs can be defined] by the number of employees and annual revenue they have. The attribute used most often is number of employees; small businesses are usually defined as organizations with fewer than 100 employees; midsize enterprises are those organizations with 100 to 999 employees. The second most popular attribute used to define the SMB market is annual revenue: small business is usually defined as organizations with less than $50 million in annual revenue; midsize enterprise is defined as organizations that make more than $50 million, but less than $1 billion in annual revenue. (http://www.midmarket.org/user-type/midsize-companies; http://www.gartner.com/it-glossary/smbs-small-and-midsize-businesses')'

For this research, 3,413 companies were selected from the group of mid-sized, mostly private, companies, using the selection criteria discussed above. (In short, the organizations selected are mid-sized companies that operate in the mid-western region of the U.S. with sales between $50 million and $1,000 million and with a number of employees between 50 and 250.) The names of the companies, their address and the top executives' names were purchased from

Dun & Bradstreet's sister company Hoover, the leading market/financial research company in the US (www.dnb.com). (Please see Appendix V for more information regarding this purchase.)

4.8.7 Pilot Study

A pilot study, or pretesting of the data collection instrument, is the last phase prior to the distribution (assuming there are no corrections that need to be made to the instrument as a result of the pretesting). Converse and Presser (1986, pg. 52) state '[t]here are no general principles of good pretesting, no systematization of practice, no consensus about expectations, and we rarely leave records for each other. How a pretest was conducted, what investigators learned from it, how they redesigned their questionnaire on the basis of it…'

As further stated by Converse and Presser (1986, pg. 52), '…the power of pretests is sometimes exaggerated and their potential often unrealized.'

For this research, the pretesting was done prior to the distribution of the questionnaire to the companies whose contact information was purchased from Dunn & Bradstreet's sister company Hoover. The pretest involved sending out the questionnaire to peers and/or current clients of the researcher that fitted the definition of being a mid-sized company located in the mid-western part of the US. The pretest administered primarily sought executives' input on the clarity of questions asked in the questionnaire.

Five pilot questionnaires were sent out in June 2013 to past or current clients of the researcher as of that time. With the exception of one returned questionnaire that asked for clarification of a single question (#65), the remaining questionnaires that were received contained comments that the questions were concise and clear and provided no suggestions for improvements. Moreover, all replies stated that the size and complexity of the questionnaire were manageable and the completion time of 30 minutes was accurate. No negative comments about the questionnaire were received. The questionnaire and the introductory letters are attached in Appendices I & X.

4.8.8 Data Collection

The collection of responses to the questionnaire was conducted via a service purchased for that purpose from SurveyMonkey (www.surveymonkey.com).

The recipients of the introductory one page letter were asked to access the questionnaire online by typing in the link to the questionnaire provided in the letter. The recipients of the letters were senior executives of organizations and, in the few cases where the name of the senior executive was not available, the letters were address to the 'President'.

Because of the fact that preparation for mailing of over three thousand letters took significant time, the data was collected in four distinct batches over the period of February 2014 to July 2014. Each batch that was sent out directed the recipient to use an individual link to the survey, thus allowing better response tracking.

The first batch corresponds to the first 1,000 questionnaires sent out in February 2014. There was a total of 10 (complete and incomplete) entries received in response to that mailing.

The second batch, which took place in March 2014, also consisted of mailing 1,000 letters. There was a total of 19 responses (complete and incomplete) generated in response to that mailing. In the third batch, 1,211 letters were mailed in May 2014 and, as a result, 21 people attempted the questionnaire (a few organizations' responses had to be discarded due to the duplication of records). The final, fourth, batch constituted the pilot cases and replies received from the business acquaintances that completed the questionnaire – these provided 9 fully completed responses.

Note that, as stated in the last paragraph of Chapter 4.5.1, with the shift of the focus of this research to a methodological direction due to the lack of an appropriate number of replies, the requirement for a firm to fit the mid-sized definition was dropped for the collection of data in batch three. As a result, two questions were added to the introductory section of the survey. The first question asked about the respondent's position in the organization and the second question asked about the industry in which the organization operated. Also, the note about the time required to complete the questionnaire was changed from 30 to 20 minutes. This change was made after seeing the average time the respondents spent on providing answers to the questions.

The responses to the questionnaire were retrieved from the collector site (SurveyMonkey) in Excel format (illustrated in Appendices II & III) and were processed according to the steps described in Chapter 5.4 and in Appendix IX.

Appendix VI provides complete details about the letter-mailing and letter-processing steps.

4.9 Data Analysis

The superiority of DM over classical statistics as a method of analysis has been expressed by a number of writers. Support for the use of DM in this research is provided next.

When considering the applications of statistics versus data mining to solve business problems, Moyar and Gardner (2012) provide an interesting, simple explanation. The writers categorize business problems into two areas: structured and unstructured. While structured problems can be solved with the use of statistics, unstructured problems are not well suited to traditional statistics, and DM's ability to interpret the characteristics and dimensions of a problem make it potentially able to generate useful contextual knowledge that could provide solutions to complex problems (Moayer & Gardner, 2012, pg. 69). Because of the facts that this research is novel and it is expected that the relationships between the independent and dependent variables will be complex, representing unstructured problem, and the number of dimensions significant, the choice of DM over statistics appears to be well justified.

Very insightful is the realization by Gullo (2015) that, due to the complexities of analysis, driven by the amount of data and the interrelationships among variables, for example, traditional data analysis techniques are no longer sufficient (2015, pg. 19). Gullo states that DM aims to fill the gaps among classical data-analysis techniques and is positioned to do so due to the fact that its interdisciplinary nature combines a number of mature fields, such as artificial intelligence, statistics, database systems and machine learning.

No work has been found to support a preference for the use of traditional statistics over data mining methods. Brusilovski and Brusilovski (2008), Fuchs et al. (2014), and Gullo (2015) claim the superiority of data mining methods that are able to capture delicate intricacies among the attributes in situations where relationships are not linear and where the data is less than perfect for the generation of positive business impact. That is not to imply, however, that traditional statistics cannot be used to extract knowledge from data.

The work of Fuchs et al. (2014) is of special interest for this research as it describes the practical application of business intelligence and analytics (BI&A) in the creation and application of knowledge for the purpose of improving business at tourist destinations, along with support for the superiority of the application of DM over traditional statistics for certain business problems. Its measurement of the intangible (tourist satisfaction) greatly adds to the importance of the paper. The work of Fuchs et al. builds on the prior published work of all authors, which serves as its theoretical foundation (Hopken et al., 2011) and it presents the practical 'knowledge destination framework' and the 'knowledge destination architecture'.

However, BI&/DM is not a magic wand or the solution to all problems, as the models and algorithms 'crunch numbers' without any understanding of the business context surrounding the numbers. Because of the importance of interpreting results and understanding of constraints, among other aspects, the involvement of experts in the DDDM appear to make it superior to classic DM approaches (Adejuwon & Mosavi (2010), Shih et al. (2010), Shollo & Galliers (2013)). Even in the context of KM, unrelated to BI&A/DM, the work of KM writers like Frappaolo (1998) calls for human involvement in the interpretation of DM results.

Because the entirety of Chapter 5 of this thesis is devoted to data, data analysis and DM models, this section only highlights the DM aspects of this research. In addition to the information contained in Chapter 5, Appendices II, III and VIII contain additional relevant information relating to DM.

A number of DM algorithms have been considered in this research. The choice of DM algorithms made and the justification thereof is presented in Section 5.5.2.

This work sets out to build an understanding about the relationships between KM and OR and how the two concepts may affect each other. As such, the DM algorithms presented in this research are Naïve Bayes, clustering, neural network and decision trees. The main difference between these algorithms is the fact that the clustering algorithm groups individual questionnaire responses into their own groups, rather than considering them as already grouped into McKenzie and van Winkelen's (2004) competence Areas.

4.9.1 DM Tool

One very recent piece of research that uses Microsoft's technologies is the work of Natek and Zwilling (2014), described in Section 3.4.4.1. While, in their research, Natek and Zwilling (2014) used what they refer to as a basic level of DM (the Excel program), this research focuses on the so-called expert level and therefore utilizes MS SQL Server as the analytical tool.

From the practitioner's viewpoint, Gartner's (2016) 'Magic Quadrant for Business Intelligence and Analytics Platforms' review states that 'Microsoft offers a competitive and expanding set of BI and analytics capabilities, packaging and pricing that appeal to Microsoft developers, independent distributors and now to business users.' The high marks achieved by the Microsoft product make it a very solid option for the analytical platform of choice for this research. While the suitability of the data mining algorithms contained in the MS SQL Server system are contrasted with the needs of this research in Chapter 6, it must be stated that the algorithms provided by the MS SQL Server platform were highly appropriate for this work. Additionally, the widely available documentation about the MS SQL Server platform, the algorithms contained in it and various on-line support communities make the platform the preferred choice for this research. Finally, the familiarity of the author of this research with the MS SQL Server platform, earned over a period of at least ten years as of the time of writing, further makes it the preferred platform.

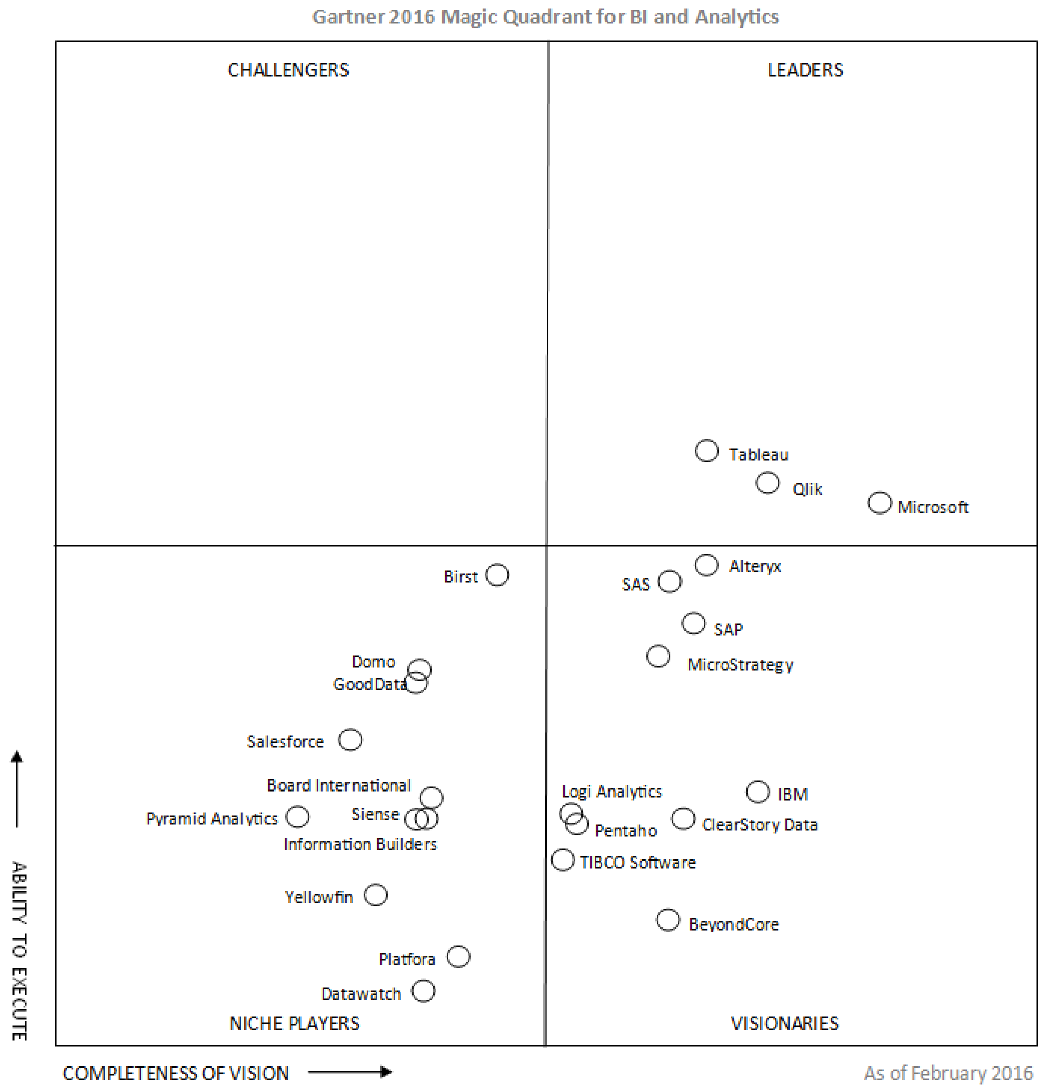


Fig. 4.9.1: Gartner's Magic Quadrant (2016) – the latest evaluation. [Derived from Gartner 2016.]

4.9.2 Summary

This chapter discussed this research from the planning perspective, providing the context for the work to be carried out. The presentation of the research was guided by the research structure presented in Section 4.2, Fig. 4.2.1.

As a result of the extensive consideration of many aspects of academic research in this chapter, this work can be stated to have the following attributes, based on the approach to classification, presented by Saunders et al. (2009, pg. 108), known as the 'research onion':

- Philosophical perspective: post-positivist
- Research type: applied
- Research approach: deductive
- Research choice: quantitative (for input data and data analysis)
- Time horizon: cross-sectional
- Methods used: questionnaire (input), data mining (analysis)

CHAPTER FIVE: CRISP-DM

5.1 Introduction

In order to govern the generation of the DM-related findings of this research, this chapter uses the industry standard CRISP-DM framework introduced in Section 2.5, as well as on the prior chapters of this work. The research work is presented in the context of each one of the six components of the CRISP-DM model, with slight adjustment to the discussion given the academic nature of this research. In Section 5.2, 'Business Understanding,' the discussion focuses on the goals of this research in the form of research questions. Section 5.3, 'Data Understanding,' describes the data used in this research, which was collected via a questionnaire. The next section, Section 5.4, 'Data Preparation,' presents the steps taken in the preparation of data for the modeling phase. The modeling phase, which details the use and workings of the selected DM algorithms, is briefly presented in Section 5.5, as the models are the subjects of their own sections in Chapter 6. Section 5.6, 'Evaluation,' examines the quality of the resulting models and their impact on prediction. The closing sections of this chapter are comprised of a short section that focuses on the deployment of the DM models and a summary in Section 5.8.

The industry standard CRISP-DM model, with corresponding chapter numbers, is re-introduced in the diagram below:

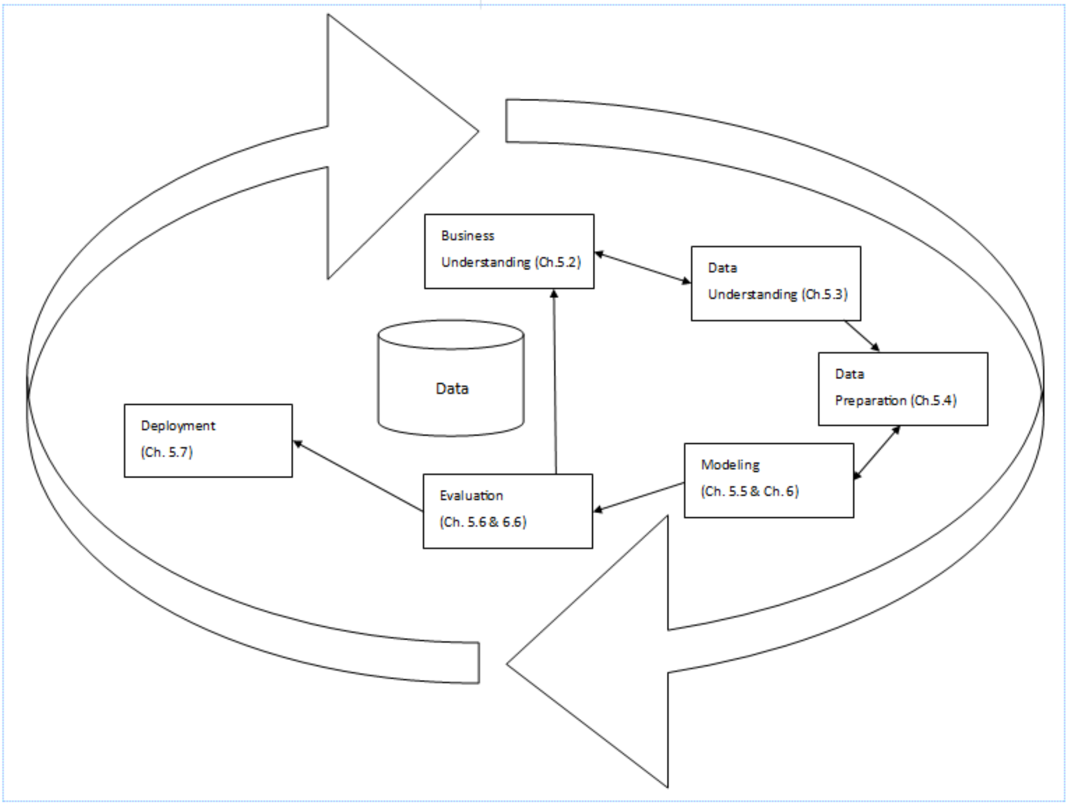


Fig. 5.1.1: CRISP-DM model. [Derived from IBM (SPSS, 2000).]

5.2 CRISP-DM: Business Understanding

According to Abbott (2014, pg. 19), the initial phase of any predictive modeling project – the definition of the project itself – is the most important part of any DM project. The reasons for great importance of project's definition within this research are numerous; some of the key factors derived from the literature review include the following:

- The need for the involvement of various types of organizational experts, such as business domain knowledge experts, data/database experts and data mining experts. Very seldom do all three fit into the mold of a single person (Cao & Zhang (2006), Brusilovsk & Brusilovski (2008), Shollo & Galliers (2013));
- Deeply affected by the point above is the need for goals and objectives for the DM project that accurately reflect business requirements (Moayer & Gardner (2012));

- An understanding of how to quantify a business objectives and the availability of data to support such quantification (Hopkins & Schadler (2015), Moayer & Gardner (2012));
- An understanding of modeling methods that can be applied to describe and/or predict business objectives, keeping in mind the DM constraints introduced earlier in this work and identified in the OR model in Section 3.3 and 3.5 (Lamot (2015), Gullo (2015), Moayer & Gardner (2012));
- A clear plan of action for the utilization of the outcomes of DM for the benefit of the organization (Hopkins & Schadler (2015), Rao (2015), Hopken (2014), Fuchs et al. (2014), Moayer & Gardner (2012), Luo et al. (2012), Ngai et al. (2009));
- An implementation plan for employing DM in operations (Abbott(2014), LeBlanc et al. (2015));
- A definition of and source of data (Larson (2012), (MacLennan et al. (2009);
- A definition of the target variable/s, if any (Larose & Larose (2015), Abbott (2014), Larson (2012)); and
- A definition of the measure of success for DM itself (Larose & Larose (2015), Larson (2012)).

When discussing the requirements and goals of the DM project with respect to this research, some preliminary relationships need to be established.

In earlier sections of this work (3.4.3.1; 3.4.3.4), it was established that, because organizational performance (OP), competitive advantage (CA) and organizational resilience (OR) are defined in similar ways by the academic researchers cited previously, treating these concepts in a similar way is justified. However, because some of the writers discuss KM's effect on OP/CA and this research considers KM processes, which can be viewed as sub-set of the field of KM, a more intimate association between KM processes and OP/CA needs to be established in order to state that KM processes (positively) affect OP. Some of the writers who explicitly discussed KM processes positively affecting OP and/or CA include Armistead (1999), Yli-Renko (2001), McKenzie and van Winkelen (2004), Ibrahim and Reid (2009), West and Noel (2009) and Chou (2011). Based on those authors' work, it can be stated that KM processes positively affect OP, and, given the similarities of the definitions of OP and OR, KM processes positively affect OR.

The process-based KM perspective utilized in this research, which was described and justified in the KM literature review (Section 3.2) and in the methodology chapter (Chapter 4), consists of six KM processes: 1) acquisition and learning, 2) storage and maintenance, 3) measurement and evaluation, 4) transfer and dissemination, 5) application and exploitation, and 6) knowledge creation. To support four out of five of the DM models used in this work, the questionnaire questions used to collect the primary data will need to be grouped into categories for the purpose of DM. Rather than arriving at a fragmented grouping based on the literature review and the classification of KM activities, the framework of McKenzie and van Winkelen (2004) is used for such groupings. To ensure that there is correspondence between the six KM processes and the competence areas used by McKenzie and van Winkelen's framework, mapping between the KM processes and the framework has been performed and is presented in Appendix IV. The fifth model used in this research (the clustering-based model presented in Section 6.3) generated its own groupings, illustrating alternative groupings of the questionnaire's answers.

Analogous to business organizations' need to define of DM goals, DM-related goals can, and must, be defined for this research. While the aims of this research have been identified and discussed in earlier sections of this work (Section 1.3, 4.6), the business goal applicable to the practical DM aspect of this research can be stated below.

In terms of the quantified results obtained from DM modeling that support this research, the end result of the DM modeling can take several forms, depending on the use and selection of algorithms (as not all algorithms provide the same output). Those include the following:

- The determination of which key KM processes impact OR, in a positive or negative way;
- The classification of an organization as resilient or not resilient;
- The identification of KM-lacking processes (inhibiting OR);
- Arriving at a score-like OR level ('OR Score'); and

- Determining if an organization is resilient.

The end results listed above are addressed in detail on a per-model basis in Chapter 6.

5.3 CRISP-DM: Data Understanding

Data understanding is the next phase after the stage of business understanding, and, as the name implies, the discussion in this phase focuses on data and data analysis. Because of its data- and measurement-heavy content, this section resembles the methodology chapter; however, because of its critical nature and its place in the CRISP-DM framework, the chapter needs to be presented on its own.

According to Abbott's (2014, pg. 20) interpretation of the CRISP-DM model, the data understating stage is used to examine and identify problems in the data, primarily to anticipate problems in the modeling phase. Janus and Misner (211, pg. 351) indicate that this CRISP-DM phase serves the purpose of pointing the analyst to the tools and/or algorithms available for the data. Similarly to Janus and Misner, Provost and Fawcett (2013, pg. 29) state that it is not uncommon for the business problem attempted to be solved by the use of DM to involve many DM tasks and that combining all of these sub-tasks into a single solution may be necessary. Data understanding is then used to identify one or more such DM tasks needed to solve the business problem.

The first analytical step in the CRISP-DM model is the data understanding phase, which, according to Abbott (2014, pg. 43) and Larose and Larose (2015, pg. 7), is used for the following purposes:

- To perform exploratory data analysis to become familiar with the data, examine key summary characteristics and individual data elements that might be masked by such summary characteristics and to discover initial insights; and
- To inspect data quality (for inaccurate or missing values, unexpected distributions and/or outliers).

Larose and Larose (2015, pg. 54) further divide the exploratory data analysis into subtasks that include the following activities:

- Examining attributes' interrelationships;
- Reaching initial insights about possible relationships between independent and dependent variables; and
- Identifying intriguing data subsets.

From the practical experience of the author of this research, one more data understanding task can be added, which is identifying (or disqualifying) the reliable target variable. This is particularly important in cases where the target variable is also used as a supervisor variable (the supervisor variable is discussed when presenting specific algorithms in Chapter 6).

Applying the guiding principles stated above, the following are the data understanding findings relevant to this research.

With the core of data understanding being summary statistics and the visualization of data, the following summaries apply to the data collected for DM analysis and used in this research. (Details of the exploratory data analysis, consisting of elements such as scatter graphs, distribution graphs and attribute interrelationship graphs, among others, are presented in Appendix III.)

Questionnaire (also referred to in this chapter as survey) data was collected between February 17, 2014 and July 25, 2014 and consisted of the collection of a total of 59 questionnaires, with 13 questionnaires being ineligible for consideration in this study. (Per discussion in Sections 4.4 and 5.4.2, incomplete questionnaires and replies from non-profit organizations were discarded; the total number of questionnaires considered was 46.) The distribution of completed questionnaires among industries is as follows:

Industry:	**Number of firms:**	**Percentage:**
Manufacturing	12	26 %
Retail	7	15 %
Construction	5	11 %
Software / Consulting / Telecommunication	5	11 %
Financial services /	3	7 %

Insurance		
Healthcare	6	13 %
Other	8	17 %

Table 5.3.1: Summary statistics about industry association of respondents

The distribution of responses with respect to organization's annual sales, number of employees and industry is as follows.

Annual sales:	**Number of firms:**	**Percentage:**
$50-$999 million (mid-sized)	46	100 %
Number of employees:	**Number of firms:**	**Percentage:**
100-999 (mid-sized)	46	100 %

Table 5.3.2: Questionnaire responses by annual sales and number of employees in organization

Finally, it is worth mentioning the practical observation of Provost and Fawcett (2013, pg. 29) relating to the data sources used in a DM project (some of those data sources can come from the outside of the organization, which requires financial investment). Provost and Fawcett (2013, pg.29) state that part of the data understanding step is estimation of the costs and benefits of each data source and the cost of processing each source to make it usable for the purposes of DM. It can be said that such aspects falls under the data constraint element of the model presented in Section 3.5: Organizational Resilience model.

5.3.1 Data Analysis

The questionnaire consisted of 84 questions; however, as discussed in the methodology section of this work (Section 4.8.3), one question (#76) was removed from the analysis. Therefore, the total number of responses to consider is as follows: 46 participants multiplied by 83 considered questions yields 3,818 individual answers that are suitable for analysis, prior to any data exclusions as a result of the analysis and the detection of outliers.

The data analyzed was retrieved from the site used to perform data collection (www.surveymonkey.com) in Excel format, as shown in Appendix II. (A more complete discussion of the preparation of the data takes place in Chapter, 5.4, which follows.) The data area of the response Excel file consisted of 84 columns, with one column per question. As mentioned in Chapter 5.4, data has been logically separated into sections: one section for each of the six competence areas, one section for assessing the performance of the organization and one for collecting the data regarding OR, forming, in total, eight sections. There are eight sections in total. As stated in Section 5.4, each answer on the Likert scale has been assigned a specific point value (N/A = 0, strongly disagree = 1, and so forth through to strongly agree = 5). The points are accumulated at the end of each section. Then, the ratio of the number of points achieved within a specific section divided by the number of possible points is computed in two formats: as a decimal number and as an integer. (Two forms of the ratio number have been calculated, a decimal and integer form, because different number formats are used in different DM algorithms, as some algorithms require an input of an integer and others require a decimal number).

Table 5.3.1.1 shows the number of questions and the number of points that it was possible to achieve within each section of the questionnaire:

Section:	**Short Notation for the Section:**	**Number of Questions:**	**Maximum Achievable Total Points:**
Create competence	CR	8	40
Exploit	EX	5	25
Decide	DE	12	60
Learn	LE	9	45
Connect	CO	12	60
Link	LI	6	30
Performance	PE	16	80
All above seven sections	7S	68	340
Organizational resilience	OR	15	75

Table 5.3.1.1: Statistics of the questionnaire sections

The summary statistics from all sections are presented in decimal form in Table 5.3.1.2, below. (The score of 1.000 in one of the ‘ratio’ columns indicates ‘Strongly agree’ answers to all questionnaire questions within that section.)

Measure:	CR:	EX:	DE:	LE:	CO:	LI:	PE:	OR:	7S:
MIN	0.475	0.52	0.3	0.222	0.4	0.267	0.338	0.333	0.389
MAX	0.925	1.0	0.933	0.978	0.933	1.0	0.92	0.96	0.899
MEAN	0.734	0.773	0.741	0.739	0.745	0.753	0.688	0.771	0.739
MODE	0.725	0.76	0.8	0.778	0.75	0.767	0.663	0.787	0.72
MEDIAN	0.738	0.76	0.758	0.744	0.75	0.767	0.688	0.78	0.738
STD. DEV.	0.1	0.119	0.117	0.132	0.11	0.143	0.124	0.116	0.096
VARIANCE	0.01	0.014	0.013	0.017	0.012	0.02	0.015	0.014	0.009
COVARIANCE	0.136	0.154	0.157	0.179	0.148	0.19	0.18	0.151	0.129
COR. COEFF.	0.56	0.239	0.665	0.592	0.656	0.484	0.761	N/A	0.711
SKEWNESS	-0.331	-00.152	-1.2	-1.234	-0.71	-1.299	-0.406	-1.22	-1.049
E.KURTOSIS	0.133	-0.531	3.34	4.0	1.163	3.044	0.421	3.225	3.250
Z-SCORE (MIN VAL.)	-2.59	-2.12	-3.78	-3.9	-3.13	-3.41	-2.83	-3.76	-3.66
Z-SCORE (MAX VAL.)	1.92	1.90	1.65	1.8	1.71	1.73	1.87	1.63	1.67

Table 5.3.1.2: Statistical analysis of each of the sections of the questionnaire

Note that the columns CR, EX, DE, LE, CO, LI and PE are the ‘independent variables’ and column ‘OR’ is the ‘dependent variable’ used in the DM models in Section 5.5 and Chapter 6.

5.3.1.1 Single-variable Summary Perspective

The mean value (0.688) for the performance section is the smallest, while the mean of exploitation (0.773) has the largest value, possibly indicating the largest number of ‘strongly disagree’ or ‘disagree’ answers to the questionnaire’s questions with the performance category. It would also appear that the exploitation category received the most favorable responses. The mean across all seven areas was 0.739.

The mode for the performance section is, again, the lowest value (0.663), which further indicates not only outliers but the majority of the answers are 'located' in the left section of the Likert scale (with the left section of the scale composed of 'strongly disagree' and 'disagree' answers). The largest mode value (0.8) is this time associated with the decide competence area. The mode value across all seven sections is 0.72.

The median value 0.688 within the performance section is once again the lowest, and the values of 0.76 for columns associated with independent variables and 0.78 for dependent variables are the highest.

The plots of individual answers against the OR section (the OR section is considered the dependent variable and all other sections are considered as independent variables) is presented in Appendix III. The plots provide descriptive confirmation of the values presented in the summary table (5.3.1.2). Visually, with the exception of the performance section, the general bulk of numbers oscillate around similar Y-axis values.

One of the properties of the normal distribution mentioned by Abbott (2014, pg. 45) is that the median, mode and mean are of the same value. Considering the values presented in Table 5.3.1.2, the values of the mean, mode and median of some of the sections (EX, CO, LI, PE) are nearly, but not exactly, the same, not clearly indicating a distribution that is close to the normal for those sections.

Another aspect of normal distribution mentioned by Abbott (2014, pg. 46) is that approximately 60% of the data will fall between the mean and +/-1 standard deviation from the mean, 95% of the data will fall within +/-2 standard deviations from the mean and 99.7% will fall within +/-3 standard deviations from the mean. Inspecting the standard deviation reported in Table 5.3.1.2 for each section, testing for the normal distribution's fit to describe the data collected leads to the following results (testing only the upper boundary, as the lower boundary is well within the limits for all sections):

Test used: Section: Mean + Std. Dev + Std. Dev + Std. Dev < MAX ?

CR:	0.734 + 0.1 + 0.1 + 0.1 = 1.034	< 0.925;	False
EX:	0.773 + 0.119 + 0.119 + 0.119 = 1.13	< 1.0;	False
DE:	0.741 + 0.117 + 0.117 + 0.117 = 1.092	< 0.933;	False
LE:	0.739 + 0.132 + 0.132 + 0.132 = 1.135	< 0.978	False
CO:	0.745 + 0.11 + 0.11 + 0.11 = 1.075	< 0.933;	False
LI:	0.753 + 0.143 + 0.143 + 0.143 = 1.182	< 1.0;	False
PE:	0.688 + 0.124 + 0.124 + 0.124 = 0.936	< 0.92;	False
OR:	0.771 + 0.116 + 0.116 + 0.116 = 1.119	< 0.96;	False
7S:	0.739 + 0.096 + 0.096 + 0.096 = 1.027	< 0.899;	False

Based on the test performed above, a normal distribution may not properly describe the spread of the data. It is, therefore, advisable to limit the use of DM algorithms that rely on the data being normally distributed.

Skewness and kurtosis are two additional important concepts applicable to data understanding and are associated with normal distribution. Abbott (2014, pg. 49) defines skewness as the measure that 'measures how balanced the distribution is'. The skewness measure for each of the sections is provided in Table 5.3.1.2. With the skewness value of 0 given for normal distribution, the value in the table shows values less than zero, indicating that all the categories reported on show negative skew. However, based on the statement made by Abbott (2014, pg. 50), only skewness with values exceeding +/-2 or +/-3 is considered significant. The significance comes into play as an effect that the skew has on the DM algorithm, calling for variable correction during the data preparation phase (Abbott, 2014, pg. 50).

Abbott (2014, pg. 51) states that 'kurtosis measures how much thinner or fatter the distribution is compared to normal distributions.' As shown in Table 5.3.1.2, E.Kurtosis values represent excess kurtosis, being the difference between the kurtosis value assumed for normal distribution (value = 3) and the value computed. Based on Abbott's discussion of kurtosis (2014, pg. 51), excess kurtosis values exceeding zero (CR, DE, LE, CO, LI, PE, OR, 7S) have platykurtic distribution and excess kurtosis less than zero (EX) have leptokurtic distribution. The graphs of the distributions confirming this statement are shown in Appendix III (figures A3.11 through A3.19).

The value of the measurement kurtosis, as was the case with skewness, becomes critical when selecting the DM algorithm, as the performance of some of the algorithms may be sub-optimal (notably, those algorithms that use standard deviation or variance in the model), requiring transformations to correct the issue.

When one considers Likert-scale responses converted to an integer and the summaries of such integers collected within each individual section as finite numbers, then the uniform distribution can be used to describe the data collected via the questionnaire. The graphs of numerical values collected within each section are presented with normal and uniform distributions in Appendix III along with the rank-order, the percentile statistics. Additionally, stem-and-leaf display of OR and 7S areas are also presented for informational purposes only, as a way to introduce an additional tool for data understanding. Finally, the other data analysis method used in this research (see Appendix III) include box plot (Fig. A3.38). From the box plot showing the range, interquartile range and the median, it can be seen that the learn ratio (LE) and the link ratio (LI) have the widest range of responses and the create ratio (CE) and the exploit ratio (EX) the narrowest (meaning a smaller range of responses).

5.3.1.2 Two-variable Summary Perspective

When looking at the association between two variables, the fact that both variables use the same units makes the measurement resistant to the weakness of the covariance measure. (Weakness is discussed by Anderson et al. [2003, pg. 108] as measuring the strength of a relationship, with non-uniform units leading to ‘greater weight’ given to the larger units.) Based on the Anderson et al. (2003, pg. 108) discussion that stated that the correlation coefficient is superior over the covariance measure when seeking to determine the linear correlation between two variables, the correlation coefficient is used in this section.

Inspecting the numeric representation of correlation coefficients presented in Table 5.3.1.2 and presented as graphs in Appendix III (Figures A3.29 through A3.36) leads to the following conclusions when relating each specific section to OR:

Section:	**Correlation Coefficient with OR:**	**Classification: (0-0.250 none, 0.251 – 0.500 weak, 0.501 - 0.750 strong, 0.751 – 1.0 very strong)**
Create competence (CE)	0.56	Strong
Exploit (EX)	0.239	None
Decide (DE)	0.665	Strong
Learn (LE)	0.592	Strong
Connect (CO)	0.656	Strong
Link (LI)	0.484	Weak
Performance (PE)	0.761	Strong
Organizational resilience (OR)	N/A	-
All above seven sections (7S)	0.711	Strong

Table 5.3.1.2.1: Correlation-related statistics for two variables

From the analysis of linear correlation, it is clearly seen that performance and OR have the strongest correlation among the variables considered. The strong linear correlation between performance and OR, therefore, appears to support the argument made in Section 3.4 very well.

5.3.2 Data Quality

This section limits the data-related analysis to the data contained in only fully completed questionnaires. All other data-related issues, including the issue of the missing values, are discussed in Section 5.4, 'Data Preparation.'

There is questionable value in summarizing points per individual company, as not all of the questions were answered by all companies, which affects the total 'points assigned' to each company's responses.

In total, there were (84 – 1) questions x 46 companies = 3,818 answers. Table 3.3.2.1, below, shows the statistics about the answers provided to the questionnaire's questions. Additional information about the individual companies' data can be found in Appendix III, Figures A3.39 – A.3.41.

Reply:	Points Assigned:	Count:	Percentage:
Not applicable	0	34	0.9 %
Strongly disagree	1	157	4.1 %
Disagree	2	549	14.4 %
Neither agree or disagree	3	570	14.9 %
Agree	4	1528	40.0 %
Strongly agree	5	980	25.7 %
	Total:	3 818	100 %

Table 5.3.2.1: Likert scale statistics

Additionally, z-score values have been computed for the MIN and MAX sum of answers within each section. (That is, the sum of the minimum and maximum ratio values has been determined for each section. Then, the z-score was computed for those sums.) In case of a z-score outside of the threshold value (discussed below), the z-score was computed for additional values that could have fallen outside of the threshold z-score value. No other z-score values were found to be outside of the threshold value besides the scores listed in Table 5.3.2.2. Computations of the additional z-scores are available in Appendix II.

The computed z-scores for each section's MIN and MAX values are presented below (with large scores, exceeding value of 3, in italic):

Section:	Z-score (Sum of MIN Score):	Z-score (Sum of MAX Score):
Create Ratio	-2.59	1.92
Exploit Ratio	-2.12	1.9
Decide Ratio	*-3.78*	1.65
Learn Ratio	*-3.9*	1.8

Connect Ratio	*-3.13*	1.71
Link Ratio	*-3.41*	1.73
Performance Ratio	-2.83	1.87
OR Ration	*-3.76*	1.63
Seven Areas Ratio	*-3.66*	1.67

Table 5.3.2.2: Detection of outliers using z-score measures

Witten et al. (2011, pg. 336) discuss various methods available for the detection of outliers 'as instances that lie beyond a distance *d* from a given percentage *p* of the training data'. The authors mention the use of statistical distribution, such as Gaussian, and fitting it to the training data and marking as outliers the instances of values with low probability. The software used in this research, MaxStat Pro 3.6, uses the Grubbs outlier test for the normal distributed data (which, according to the software's on-screen hint, would require assurance that the data can be reasonably approximated by a normal distribution through the Anderson-Darling test). Finally, the standardized values (z-scores) are a well-known measure for the location of outliers for a bell-shaped distribution (which applies to the data used in this research – per distribution graphs in Appendix III) according to Anderson et al. (2003, pg. 97). Anderson et al. (2003, pg. 97) state (because all of the data will be within +/- 3 standard deviations of the mean), '[h]ence, in using z-scores to identify outliers, we recommend treating any data value with a z-score less than -3 or greater than +3 as an outlier'.

As suggested by Anderson et al. (2003), the standardized z-score value has been selected in this research as the method of identification of outliers, especially since is possible to calculate such value using the widely adopted Excel environment and the focus of this research is not on specific output in numerical form.

Based on the data presented in Table 5.3.2.2, it is apparent that six out of nine (66.7%) z-score values exceed the value of -3.0. The data plot presented in Appendix III (Fig. A3.37) visually presents the location of outliers.

Note that, within the link ratio of Fig. A3.37 mentioned above, it appears that there are two 'low values' that both possibly exceed z-score values of -3. It has been determined that, while the lower point on the graph has the z-score value of -3.41, the point immediately above it has the z-score value of -2.94, which is not less than -3.0.

According to Abbott (2014, pg. 86), there are a number of approaches to outliers. The most common methods of dealing with outliers include the following:

- Removal of outliers from the modeling data;
- Separation of outliers and the creation of a model specifically for the outliers;
- Transformation of the outliers so they are no longer outliers;
- Binning (conversion to categorical type) of the data; and/or
- Leaving the outliers in as part of the modeling data.

Because of the nature of this work, which emphasizes the process and its feasibility over the specific and actionable outcome produced by the DM, the outliers (one company's answers) have been removed from the modeling data.

5.3.2.1 Data Audit

Examining trends and identifying problems in the data and visualizing the data fall into the area of data audit.

Based on several suggestions made by Abbott (2014, pg. 81) about the data understanding phase in DM modeling, the following remarks about the modeling data used in this research can be made:

- There were no missing values in the responses considered in this chapter, as only completed questionnaires were considered. As stated in Section 5.4, only questionnaires that were answered completely were considered in the analysis;
- The maximum values all had a z-score below +2. There was one response to the questionnaire that had to be discarded due to the excessive number (66.7 %) of outliers. The number of unique companies that provided modeling data is therefore 46 – 1 discarded entry = 45 companies;
- For algorithms assuming normal distribution of data, the largest skew (of -1.8) is reported by a question in the OR section. There are few individual answers with values of 0. The sections fall within the

following range of skew: -1.299 to -0.151. (Sizable skew can make some DM algorithms unusable, given the data set);

- Kurtosis (information for algorithms affected by excessive kurtosis) for individual questions varies widely between -1.8 and 4.3. For all sections the range is between -0.531 to 4.0. (Large kurtosis can make some of the algorithms unusable given the data set);
- There are no responses with a predominately single response to all questions; and
- There exists a relatively strong correlation (correlation coefficient = 0.761, where 1.0 indicates a perfect correlation between variables) between performance and OR (which supports the argument presented in Section 3.4.3).

5.4 CRISP-DM: Data Preparation

5.4.1 Background

‘Real-world data is dirty. Often you’ll have to do some work on it before you can use it’ (Grus, 2015, pg. 127).

While the data to be used in a model can present a very large number of unique problems, the primary issues addressed here are those applicable to this research (looked at from a broader view). However, for the sake of completeness, some of the most critical issues encountered in the data preparation phase are also briefly mentioned.

The preparation of the input data, which takes between 60 and 90 percent of the time of the entire predictive modeling project (Abbott, 2014, pg. 83), either follows, or can be carried out simultaneously with, the data processing. The goal of this phase is to convert input for modeling data into a form that is better suited for a particular DM algorithm. While this research uses a single-formatted input data set, a typical commercial application involves the use of numerous data sources generated by different systems, each with their ‘own’ data problems. The discussion here regarding data preparation goes slightly beyond the needs of this research in order to illustrate issues that may need to be addressed in commercial research of a similar type.

Abbott (2014) divides his discussion of the data preparation phase into discussions of variable cleaning and feature creation; he goes on to describe numerous approaches to data preparation within each one of the two main tasks.

The following tasks are mentioned by Larose and Larose (2015, pg. 8) as involved in the preparation of data: the selection of variables for analysis, the transformation of variables (to achieve normality of data), if needed, and data cleaning.

Han et al. (2012, pg. 84) view data preparation from a data quality perspective, stating that data needs to satisfy requirements for the intended use. The factors affecting data quality identified by Han et al. (2012, pg. 84) include accuracy, completeness and consistency. Other factors mentioned, but that do not necessarily affect data quality, include timeliness (meaning that data are received on a timely basis), believability (others trust the data) and interpretability (ease of understanding the data). The tasks involved in data preparation, according to Han et al. (2012, pg. 85), include the following:

- Data cleaning – resolving missing values, smoothing noisy data, identifying or removing outliers, and/or resolving inconsistencies;
- Data integration – integrating multiple data sources (databases, Excel files, text files and so forth);
- Data reduction – reducing the representation of the data volume, which includes the following:
    - Dimensionality reduction – obtaining reduced ('compressed') representation of the original data; and
    - Numerosity reduction – replacing the data using a smaller representation of the data.

Witten et al. (2011, pg. 60), in addition to the similar points about data preparation noted in the discussion above, emphasize the importance of involving domain experts in addressing data-related issues so that appropriate assumptions about the data can be made.

Foster and Fawcett (2013, pg. 30) mention an important concern with regards to data preparation, which is the concept of 'leaks', stating that '[a] leak is a situation where a variable collected in historical data gives information on the target variable – information that appears in historical data but is not actually available when the decision has to be made.'

Some examples of real-life data preparation tasks include converting data to a tabular format, removing or inferring missing values and converting data to a different data type.

For the purpose of this research, based on the work of the authors quoted in this chapter as well as the industry experience of the author of this thesis, the discussion of data preparation includes the following:

- Data cleaning
    - Handling missing/incorrect data
    - Identifying misclassifications of categorical variables

- Data transformation
    - MIN-MAX normalization
    - Z-Score standardization
    - Decimal scaling
    - Transformations to achieve normality
    - Flag variables
    - Transforming categorical variables into numerical variables
    - Discretizing numerical variables
    - Adding an index field
    - Removal of unneeded variables
    - Removal of duplicate records

5.4.2 Preparatory Steps

Prior to the discussing the specific steps involved in data cleaning and transformation, a short discussion about the data itself is necessary; the preparation of the instrument was discussed in Chapter 4.

- The questionnaires collecting data were located at www.surveymonkey.com, a site specifically designed to administer and manage surveys. The data was collected in four batches:

- Batch # 1: from a questionnaire constructed on January 25, 2014: 10 records.
- Batch # 2: from a questionnaire updated on March 5, 2014: 19 records.
- Batch # 3: from a questionnaire updated on May 21, 2014: 21 records.
- Batch # 4: holds the pilot cases introduced into the system records: 9

Total number of input records: 59

Out of the 59 collected replies (containing 84 questions):

- Five replies answered 0 questions.
- Two replies answered the first 7 questions.
- One reply answered the first 8 questions.

Total number of incomplete answers records: 8

Because the respondents of the eight incomplete questionnaires terminated the questionnaires very early on, those eight responses were eliminated from the input data set. According to Larose and Larose (2015, pg. 23), disregarding entries with missing values is a common practice. (Note, per the discussion of the choice of companies that took place in the methodology section of this work, five replies were been eliminated due to the fact that they came from educational institutions and not private, mid-sized companies operating in a non-academic industry.)

The exclusion mentioned above, along with the exclusion of the outlier identified in Chapter 5.3, yields the following total of input data:

59 (total responses) – 8 (incomplete) – 5 (academic) – 1 (outlier) = 45 (used).

The discussion that follows relates to the 45 records that compose the input data set. (This makes the total number of answered questions 3,735, and the total number of utilized answers to 45 x 83 = 3,735. Per the discussion in Chapter 4, question #76 was discarded. The discussion that follows relates to the remaining 83 questions.)

The four batches collected from the survey administering site, each of which was collected in an individual Excel file, were combined into one master Excel file. Later, this combined file was stored in a table created for that purpose. This table, collecting all valid (45) responses, was created in Microsoft SQL Server 2012.

The steps of data preparation typically begin with the careful analysis of the data to be used in modelling with a goal of identifying all data anomalies.

5.4.2.1 Data Cleaning

According to Larose and Larose (2015, pg. 20), the most common problems calling for careful data cleaning are as follows:

- Obsolete or redundant fields;
- Missing values;
- Outliers;
- Format of data not suitable for DM algorithm/s; and
- Out-of-the-ordinary values (i.e. values that are not aligned with common sense).

The outliers were addressed in Section 5.3 on a 'per section' level. Outliers on a 'per-question' level have not been determined, as, given the composition of the organizations studied, a wide range of response is expected, including the 'N/A' response that receives the numeric value of zero, which greatly affects the identification of outliers. As long as all of the questions within a given section fell within +/-3 z-scores, each question within a given section was accepted. Finally, while the individual answers to the administered instrument are important to this research, its main focus is to show the applicability of the DM methods, not the interpretation of the results obtained from the application of DM.

Variable cleaning refers to the correction of the variable itself. For the purpose of this work, the variable will represent each one of the questions (which are represented in the Excel input data file as a single column).

All of the variables used in the research are of the categorical types, which are later converted into equivalent integer (finite) value. As suggested by Abbott

(2014, pg. 84), incorrect values of categorical variables are very difficult to uncover and, typically, graphical methods of data presentation are used in order to inspect these types of variables.

To overcome the dependency on extreme values (in the case of this research, the 'N/A' answers), the interquartile range (IQR) is used as a measure of variability in the discussion of the individual variables used in this research. (It is understood that IQR represents the difference between the 3rd quartile and the 1st quartile, meaning that IQR is the range for the middle 50% of the data. A box plot is used to provide a graphical summary for each variable.) According to Hartwig and Dearing (1979, pg. 23), 'the box-and-whisker provides detail when it is often needed most, whenever one or both of the tails of a distribution contain extremely large or small values.'

The box plots based on the IQR, the low limit (the smallest numeric value of the answer, transformed into a number value), the upper limit (the largest value of the answer, transformed into a number value) and the median for each variable are presented in Appendix III (Figures A3.42 to A3.49). Descriptive statistics corresponding to these graphs are included in Appendix III, in Figures A.3.50 to A.3.57. Appendix I contains the survey questions discussed in this section. Finally, results are generated using the MaxStat Pro 3.6, Easy Fit 5.6 and Excel software.

The Create section of the questionnaire made inquiries about knowledge creation and acquisition within an organization. Eight questions from the Create section of the questionnaire were considered. All questions were answered, so no method for solving missing data was necessary. Five questions, Create_GapId, Create_GapFix, Crete_GapSatisfy, Create_Facilities and Create_Insight, had the smallest IQR, with Create_Insight also having the smallest range of answers (there were no answers in the 'disagree' or 'strongly disagree' range), meaning that every respondent had been engaged in insight generation. The Create_GapSatisfy answer did not generate a single 'strongly agree' reply when asking about the knowledge gap and the extent to which such a gap was being addressed at the organization. The two questions with the largest IQR are Create_Employees and Create_Suggest, perhaps indicating slightly larger disparity concerning the importance of allowing employees to reflect on their jobs and to record and store employees' suggestions about improvements related to their jobs. The highest mean, 4.5, was recorded in the Create_Insight question and the lowest, 2.2, in the Create_GapSatisfy question.

The Exploit section asked questions related to the exploitation of existing knowledge within an organization. There were five questions in this section. Similarly to the previous section, there were no missing answers. (Since there were no missing data in all of the 46 replies that comprised the input data, the discussion of missing data is omitted.) Four questions had an IQR of one point on the scale: Exploit_References, Exploit_Simulate, Exploit_Consult and Exploit_Reflect. The questions Exploit_Consult and Exploit_Reflect did not receive any 'strongly disagree' answers, with the answers to the question Exploit_Reflect being uniformly distributed (50% of answers between 'neither agree nor disagee' and 'agree' and 25% between 'agree' and 'strongly agree,' as well as 25% between 'N/A' and 'disagree'. Most favorable answers (primarily 'agree' and 'strongly agree') were received when firms were asked about referring work and seeking internal consultation prior to the undertaking of a major project. The highest mean of 4.2 was recorded by the answer to the Exploit_References question and the lowest mean of 3.5 by the answer to the Exploit_Simulate question.

The Decide section asked questions related to decision-making and decision alignment (with strategy) within an organization. This section had twelve questions. Replies from 2 out of 46 responses (to the questions Decide_Alliances and Decide_Intelligence) contained 'N/A' responses. The majority of 'agree' and 'strongly agree' responses were given to the questions that asked if an organization forms alliances and joint ventures (Decide_Partnership), if they view professional organizations as learning opportunities (Decide_Professional), provided work conditions (Decide_Condition) and set boundaries for decision-making (Decide_Boundaries). The replies to the question about the use of CRM as a strategic tool (Decide_CRM) appears to have the largest IQR. The question about sponsoring and/or supporting academic research (Decide_Academic) received mostly neutral and 'disagree' answers. The highest mean of 4.2 was recorded by the Decide_Partnership, Decide_Professional, Decide_Condition and Decide_Boundaries questions. The lowest mean of 2.5 was recorded by the Decide_Academic question.

The Learn section investigated individual and organizational learning, asking nine questions. One out of 46 organizations replied 'N/A' to the question Learn_Venue within this section. Four questions received primarily 'agree' and 'strongly agree' replies: Learn_Training, Learn_Mentor, Learn_Reimburse and Learn_Portal. The reply to the question asking about learning taking place with the help of data mining (Learn_BI) generated the widest range of answers, from 'disagree' to 'strongly agree', indicating, for the purpose of this research, mixed utilization of DM within organizations. Perhaps most surprising was the relatively low 'score' in the area of capturing lessons learned (Learn_Capture). The largest mean, 4.2, was associated with the answers to the Learn_Training question. The lowest mean of 3.1 was associated with answers to the Learn_Capture question, and the 1.4 standard deviation of the Learn_BI was the largest reported.

The Connect section posed questions related to the connecting of intra-organizational activities with activities occurring outside of organizational boundaries. The section contains twelve questions. Out of 46 companies, one firm chose the 'N/A' answer to the Connect_Buying question. From the group of questions in this section, the question dealing with building customer relationships (Connect_Relations) received the most 'agree' and 'strongly agree' replies. Other highly 'scored' answers included responses to the following questions: Connect_Beliefs, Connect_Confident and Connect_Educate. The questions with the largest IQR included questions about vendor coalition and education (Connect_Buying), connecting with firms in other industries (Connect_Activities) and cooperation with competitors in the areas outside of competition (Connect_Resources). The highest mean was reported by the Connect_Relations (4.5) question and the lowest (3.0) by the Connect_Resources question. The standard deviation of 1.3 associated with the Connect_Buying question was the largest reported within the Connect section.

The Link section examined the existing business links of an organization. The section contained six questions. The 'N/A' answer was recorded as an response at least once to all, except the Link_Relationship question, which had no 'N/A' responses. Interestingly, the answers to the questions, except the answer to the Link_Relationship question, were relatively similar in terms of IQR and standard deviation (1.1 – 1.2). The question asking about whether the organization had recently formed relationships with customers, suppliers and external partners (Link_Relationship) had a majority of 'agree' and 'strongly agree' answers. The highest mean of 4.3 was also reported with an association to the Link_Relationship question. The smallest mean of 3.2 was reported by the Link_Leadership question.

The Performance section attempted to evaluate the current as well as future performance of an organization. This section also served another purpose: to validate the argument made in Chapter 3 that relies on the correlation of organizational performance with OR. The Performance section had sixteen questions. Five questions from this section recorded an 'N/A' answer at least once. In general, the majority of the replies had an IQR of two or more, fluctuating between the 'disagree' and 'agree' answers. (The questions Performance_Financial, which inquired about financial gains from activities that improve products/processes and Performance_Copyright, which focused on trademarks and copyrights obtained, had a much larger IQR, indicating significant fluctuations with regards to these activities at various firms.) Of interest are the answers to the Performance_Strategy question asking about challenged business strategy and to the Performance_Problem question that dealt with view problems in a constructive way. The answers to the Performance_Strategy question were primarily in the 'agree' and 'strongly agree' range, whereas, in the case of the Performance_Problem question, the answers had an IQR of zero and median at the 'agree mark', indicating that most of the organizations view (or attempt to view) challenges in a positive way. The lowest mean of 2.5 was recorded within the answers to the Performance_Copyright question and the largest, of 3.8, within the answers to the Performance_Strategy question.

The OR section attempted to measure the level of OR within an organization. Fifteen questions were considered in this section, as question #76 (which addressed turnaround), as previously discussed, was removed from consideration. Within the results, there were at least five 'N/A' answers, primarily in response to the questions about the financial and market-share performance of an organization. The median of 5 for the answers to the questions OR_Income10, OR_Income5, OR_Assets10 and OR_Assets5 indicates improvement in financial conditions since the financial crisis of 2008. Other questions that had responses with box plots located at the top of the scale (at 'strongly agree') included OR_External, OR_Tolerance and OR_Change. Responses to the question (OR_Denial) asking about denial-free, arrogance-free and nostalgia-free responses to changes in business conditions generated the lowest 'score' and had the largest IQR. The median for the answers to this question was 2, and the mean, also the lowest for the section, was 2.7. The highest mean of 4.4 was reported for the OR_Asset question.

5.4.2.2 Data Transformation

Larose and Larose (2015, pg. 30), while discussing the differences that various data ranges have on data mining algorithms, state 'data miners should normalize their numeric variables, in order to standardize the scale of effect each variable has on results.' Han et al. (2012, pg. 113) indicate the importance of data transformations by normalization: 'Normalizing the data attempts to give all attributes an equal weight. Normalization is particularly useful for classification algorithms involving neural networks or distance measurements such as nearest-neighbor classification and clustering'. The discussion below is based on the material contained in the books of Han et al. (2012) and Larose and Larose (2015) and on methods applied in the industry.

With regards to this research, only limited data transformations took place. Despite that, to ensure the completeness of this work, especially when referenced by a practicing professional, a short description of the most common data transformations is provided next.

- MIN_MAX normalization: MIN_MAX normalization works by determining how much greater the field value X is than the minimum value min(X) and scaling the difference by the range. The values for this normalization range from 0 to 1, with the MIN value of X assuming the normalized value of 0 and the MAX value of X assuming the value of 1.
- Z-Score standardization: This method works by taking the difference between the field value and the field mean value and scaling the difference by the standard deviation of the field values.
- Decimal scaling: This method ensures that every normalized value lies between -1 and 1. (It does so by dividing the number X by the 10 to the

power of *d,* where *d* represents the number of digits present in number X.)

- Transformations to achieve normality (in order for a variable to resemble normal distribution): The goal of this method is to achieve symmetry and normality in the distribution of a variable. To eliminate skewness, transformation to the data (log, square root transformation and perhaps also inverse square root transformation) is applied. (Sometimes, experimentation with further transformations is also needed in order to yield normality.)
- Flag variables: 'A flag variable (or dummy variable, or indicator variable) is a categorical variable taking only two values, 0 and 1' (Larose & Larose, 2015, pg. 39). Variables taking only the values of 0 and 1 (also referred to as binary variables) are often used to designate presence or absence of some sort. For example, 0 may indicate an organization that is not resilient and 1 a resilient one.
- Transforming a categorical variable into a numerical one: As stated by Larose and Larose (2015, pg. 40), this type of transformation is typically to be avoided, as it introduces a certain order that might not hold in reality, with the exception being survey responses. This type of transformation has been used in this research, as each response was assigned a numerical equivalent ('N/A' = 0, 'strongly disagree' = 1, 'disagree' = 2, 'neither agree nor disagree' = 3, 'agree' = 4, 'strongly agree' = 5). The responses and the assigned equivalent numerical values are clearly ordered.
- Discretizing numerical variables: This is a very common method of providing input data that is of a continuous type into a DM algorithm that expects discrete values. Essentially, the numbers constituting an input set are divided (discretized) into buckets. In this research, numerous trial models have been constructed using this concept in order to determine the differences in output results.
- Adding an index field: This is a very common requirement in almost all DM algorithms; it is used to track the order in the table and to identify each individual record in a table, among other things. The data structure used in this research uses an index field (named IP).

- Removal of unneeded variables: While, in the industry, the removal of any kind of variables is typically highly discouraged in the practical application of DM, some variables do not provide any value to the model and can be removed. Larose and Larose (2015, pg. 43) state that unary variables (those that take only a single value) and nearly unary variables can be removed in order to reduce storage space, model size and model processing requirements.
- Removal of duplicate records: Being one of the easier issues to spot, duplicate records should be removed from the input data set to ensure that they play no role in prediction and to reduce space and processing requirements.

Some of the normalization methods, if required by a specific DM algorithm, will be discussed further in the section that describes the uses and the outcomes of specific DM algorithms.

5.4.2.3 Application of the Data Preparation Phase to this Research

Data preparation consists of a set of highly complex tasks. Some of the common tasks that are performed during data preparation, in addition to the short list of tasks mentioned at the beginning of Section 5.4.2.1, include the following:

- Simple variable transformations;
- Fixing skew;
- Binning (discretizing) continuous variables; and
- Variable selection (prior to modeling)

Finally, the nature of the research, with its focus on the methods rather than on numerical results, did not justify carrying out skew fixing and some of the variable transformation tasks due to the very limited value to be gained. Moreover, the data preparation phase is typically not entirely completed on the first attempt. As stated by Abbott (2014, pg. 143) '[d]o not consider Data Preparation a process that concludes after the first pass. This stage is often revisited once problems or deficiencies are discovered while building models.'

5.5 CRISP-DM: Modeling

Each DM algorithm is the subject of its own section of this thesis, and each of these sections includes a discussion of the model's requirements and construction, as well as the findings. The models have therefore also each been given their own section, one per DM algorithm type with Section 6.2 discussing two Naïve Bayes models, Section 6.3 the clustering model, Section 6.4 the neural network model and Section 6.5 the decision trees model.

The purpose of this section is first to provide DM-based information common to all DM algorithms before moving on to the next sections, which complete the discussion of CRISP-DM.

5.5.1 Technical Information

The DM component of this research is based solely on Microsoft's platform. Microsoft Visual Studio 2010 (Version: 10.0.40219.1 SP1Rel) is used as the tool for DM model building, processing and interpretation. Microsoft SQL Server 2012 (Version: 11.0.5343.0) is used as the back-end database.

For the purposes of data storage and building DM models, a database called 'RGU' was created. A table holding the survey's data, located within the 'RGU' database, has been called 'tbl_DM_KM_OR_RGU'. This database and table were used by all the DM models explored in this research. (The definition of the tbl_DM_KM_OR_RGU table can be found in Appendix VIII, Fig. A8.1. Fig. A8.2 illustrates the location of the 'RGU' database, the 'tbl_DM_KM_OR_RGU' table and the small data content of the table as it appears in Microsoft SQL Server Management Studio.)

The following components of the DM/BI environment are common to all of the models presented in this chapter:

- Microsoft Visual Studio 2010 development environment. The solution, the highest hierarchy level in the development environment, is called 'RGU_Project'. The solution 'RGU_ Project' consists of two main parts: the data load part, called LoadTestData_RGU and the data mining part, called 'RGU'. (The data load part responsible for the loading of the data from the Excel file retrieved from the questionnaire collector website into the database will not be discussed here but is described in Appendix IX.) All of these concepts are illustrated in Appendix VIII, Fig. A8.3.

- The data source 'RGU_Analytics' represents a connection to the MS SQL Server's database (also called 'RGU'). Fig. A8.4, in Appendix VIII, provides a pictorial illustration of this concept.
- The data view 'RGU_DInfSc' provides additional granularity of data access, granularity to the table and table field level (and this research uses only a single database and a single table). This concept is presented pictorially in Fig. A8.5, in Appendix VIII. Data source and data source view, in fact, provide the interface to the data residing in the database.

Two additional key components common to all DM models of the data mining project based on the MS SQL Server platform are the mining structures and the mining models. As stated by MacLennan et al. (2009, pg. 148) '[a] mining structure defines the domain of a mining problem, whereas a mining model is the application of a mining algorithm to the data in a mining structure.' That is, the mining structure refers to the information about the data available to be used in data mining, such as the list of the table's columns, each column's data type and optional flags. In addition, the mining structure contains a list of DM algorithms that can operate on the data from the mining structure. The mining model contains the DM algorithm, any parameters passed to the algorithm and a list of columns from the mining structure. Because different DM algorithms can use different elements of the mining structure and can require different parameters, the mining structure and mining models are described further when considering specific DM models. (For the purposes of the discussion in this section and this research as a whole, when referring to the DM model, unless otherwise stated, the reference will also include its underlying mining structure.)

5.5.2 Data Mining Models

The following sections of this chapter present the applicability of specific DM algorithms for measuring the impact of KM on OR as well as seeking to satisfy the aims and objectives of this research as stated in Chapter 1.

The MS SQL Server 2012 tool used in this research supports the following DM tasks, or techniques (task and technique being used interchangeably), ordered alphabetically. Appendix VIII, Fig. A8.8 illustrates the list of mining techniques available in MS SQL Server 2012.

DM task/technique:	**Reason for selection in this research:**	**Reason for not selecting the task in this research:**
Association rules		Association, also known as 'market basket analysis,' attempts to find patterns in a group membership: which items occur together and which items can be added to the group? While the association rules method may offer significant findings in relation to which KM processes occur together in the resilient organization, such a determination is beyond the scope of this work, as its focus is on the primary (singular) KM processes that contribute the most to the OR of a firm.
Clustering	The clustering technique attempts to find natural groupings of KM processes within resilient organizations, directly supporting the	

	aims and objectives of this research. Moreover, the clustering method can arrive at natural groupings (not necessarily mapping onto the competence areas) of KM activities responsible for OR.	
Decision trees (*)	This research utilizes the decision trees algorithm to investigate the knowledge that can be generated by an algorithm using the 'if-then-else' construct it generates. In addition, the algorithm is used in an attempt to answer the question of what makes an organization resilient? Additionally, (based on the tree splits) this algorithm is used to seek answers to the question of which KM processes are the most influential in determining the OR of an organization. The use of the algorithm is also expected to support the quantification of results mentioned in Section	

	5.2.	
Linear regression		This research does not seek to determine a linear relationship between two numeric variables or to find the patterns that describe numerical values.
Logistic regression		As a special (simpler, one-layer) case of neural network that models 'true/false' outcomes, supporting at best the quantification of results mentioned in Section 5.2 and not the research question itself, this method is not pursued in this work. Instead, a neural network (consisting of one or more layers) and decision trees (supporting 'true/false' types of predictions) are used.
Naïve Bayes	Seen by many writers as well as practitioners as the starting point for predictive analysis, this method is employed to better understand the input data and its relationship with the	

	competence areas as well as its impact on OR. The method will assist in answering the question of what relationships currently exist between KM processes and OR? (Two models have been constructed: one for the purpose of investigating the attribute relationships and one for predictive modeling purposes.)	
Neural network	This technique is used for the purpose of classification, so the notes related to Naïve Bayes apply to this technique as well. (That is, DM models will attempt to illustrate the use of NN for the purpose of determining which KM processes are the most influential on OR, taking advantage of the ability of the algorithm to determine complex relationships among the data.) In addition, the use of the algorithm supports the quantification of results	

	and goals mentioned in Section 5.2.	
Sequence clustering		This research does not seek to find patterns in a series of events (such as a series of KM processes that take place at an organization). The focus of this research is on a single-level of KM processes and their relationships, instead of the analysis of KM processes that occur in a sequence.
Time series		With the 'time factor' not being of significant importance in this research, forecasting any future numerical values is, clearly, not of any interest.

Table 5.5.2.1: SQL Server 2012-based DM Algorithms

(*) MacLennan et al. (2009, pg. 236) state that the algorithm is called 'trees' instead of 'tree' because of the possibility of building different trees based on parameters and splitting criteria as well as the possibility of creating multiple trees targeting multiple attributes in a single model. This is illustrated in Section 6.5.

Based on the information in Table 5.5.2.1, the DM techniques selected for this research (with comments in the center column) are clustering, decision tree, Naïve Bayes and neural network.

5.6 CRISP-DM: Evaluation

This chapter examines methods of evaluating DM models prior to their release into the production (or everyday use) environment. The evaluation stage allows for evaluation of the results of the DM models, which may perhaps necessitate additional model changes, as well as comparison of the results of the models (where applicable).

This chapter discusses general issues affecting the quality (referring to the ability to reflect reality) and performance of the DM models; it looks at the tools that are part of the development platform. Such tools include accuracy charts, classification matrices and cross-validation. Illustrations that support the discussion can be found in Appendix VIII.

5.6.1 General Information

The construction of the DM models is, as indicated by the CRISP-DM model itself, is highly iterative process, many times requiring multiple attempts at each stage in order to arrive at the final DM model. As was already presented, the CRISP-DM methodology is complex, providing opportunities for many challenges that affect DM models to arise. Some of the DM challenges have been discussed previously, but the most common ones are re-stated here:

- Data – missing or inaccurate data, correlated variables, sample size and similar issues;
- Data mining tool – selecting the proper algorithm, setting up the tool's parameters, etc.;
- Usability – ensuring that the resulting model addresses the original goals and works outside of the development/testing environment;
- Nature of the problem – not allowing the focus on the technical details for the model to answer the wrong question; and/or
- Modeler's skills – often, unqualified people are given responsibilities as of data miners/data scientists, just so that the organization can 'jump on the DM bandwagon.'

From the professional experience of the author of this work, there is still a great deal of skepticism in the field with regards to the application of DM models in real-life situations, especially those that have profound effects on a business. It is therefore imperative that the resulting models be given extensive scrutiny in order to reduce such anxiety and, in the end, do more good than harm.

Provost and Fawcett (2013, pg. 31) state that '[t]he purpose of the evaluation is to assess the data mining results rigorously and to gain confidence that they are valid and reliable before moving on. If we look hard enough at any dataset we will find patterns, but they may not survive careful scrutiny.'

On the highly technical side, Larose and Larose (2015, pg. 452) describe model evaluation according to the three most common modeling techniques: descriptive modeling, estimation and prediction and classification tasks.

With regard to descriptive modeling, the authors simply state that the best representation/description is the one that 'minimizes the information required (of bits) to encode (i) the model and (ii) the exceptions to the model.'

For the estimation and prediction techniques, where the estimate and predicted values are known, Larose and Larose suggest the use of the mean square error and mean absolute error functions, represented by a mathematical formula.

The model evaluations that measure for the classification tasks (the majority of tasks presented in this research) involve the following evaluative concepts. For the binary classification in the discussion below, the following outcomes of classification are assumed:

	Prediction	|	Outcome
True Positive	T	|	T
True Negative	T	|	F
False Positive	F	|	T
False Negative	F	|	F)

- Model accuracy – refers to computing the overall measure of the proportions of correct classifications.
- Overall error rate – similar to the above, but computes the proportions of incorrect classification.

- Sensitivity and specificity – (for binary classifications) sensitivity measures the ability of the model to classify a record positively. Specificity measures the ability of the model to classify a record negatively.
- False positive and false-negative and expressed as: (rate – (for binary classifications) false positive is an inverse of specificity (Equal to 1 – specificity). False negative is an inverse of sensitivity (1 – sensitivity).
- Proportions of true positive and true negatives – (for binary classifications) – as ratios of true positives divided by the sum of false positives and true positives and true negatives as ratios true negatives divided by the sum of false and true negatives.
- Proportions of false positives and false negatives – uses similar concept to the one described above.
- Misclassifications costs – impact of misclassifications: false positive, false negative and adjustment necessary on the performance of the algorithm.
- Cost-benefit table – table based cost vs. benefits analysis, comparing all four of the possible classifications.
- Lift and gain charts – graphical representation of assessing and comparing the usefulness of classification model.

While the testing techniques presented by Larose and Larose (2015) provide better assurance of model correctness, only the methods supported by the SQL Server Visual Studio development environment are discussed in this work. The remaining methods have been provided to ensure the completeness of the discussion of the testing and validation of DM models.

Two of the frequently encountered issues in DM techniques are generalization and overfitting (or the inability to perform these). The concept of generalization is directly related to the issue of overfitting: 'Generalization is the property of a model or modeling process, whereby the model applies to data that were not used to build the model (Provost & Fawcett, 2013, pg. 112)'; that is, the model fails to generalize beyond the training data that it has already encountered. And, as a reminder, overfitting is a process that occurs at the expense of generalization, as it is a result of the tendency of DM algorithms to tailor to the training data and not the general population.

5.6.2 DM Model Evaluation

In order to ensure a logical flow to the content of the material presented in the thesis, the evaluation of the DM models is presented in Section 6.6 following the presentation and discussion of the DM models.

5.7 CRISP-DM: Model Deployment

While the deployment of the developed model is outside the scope of this academic work, as the models presented in this thesis were developed with the single purpose of supporting this research, a brief summary is included to ensure that the discussion of the CRISP-DM model is complete. The summary illustrates some of ways models are 'consumed' by their users.

De Ville (2001, pg. 51) states that '[t]he main task in deployment is to create a seamless process between the discovery of useful information and its application in the enterprise.' By "seamless," the author implies that the knowledge generated by the application of the DM models should be released to the wider public in easily usable form. The deployment of the DM models should take place after proper model testing and validation using the techniques discussed in the previous chapter. Moreover, thought should also be given to the maintenance (including re-validation) of the models, as they will need to be regenerated occasionally to keep them current. Lastly, deployment will vary greatly depending upon real-time or near real-time knowledge presentation requirements.

While the descriptive report mentioned by de Ville (2001, pg. 51) still prevails in the field, Janus and Misner (2011, pg. 361) highlight several key aspects of and options for deployment:

- The use of SQL Server's Integration Services, which allow for the automation of delivery of the knowledge created by DM models (including the automation of preparatory steps such as data loading and model processing);
- The inclusion of the DM models into on-line analytical processing (OLAP) (into multidimensional models as one of the dimensions of the data cube itself). Later, such added dimensions can facilitate the

analysis of groupings and trends discovered by the model (this approach of adding the DM model as a cube's dimension will only work for decision tree, clustering or association DM methods);

- With the help of the DMXs (data mining extensions) embedded in SQL Server, the DM models can be made accessible to the Reporting Services, a layer in Microsoft's product range that is part of SQL Server 2012 and allows the creation of parametrized reports;
- Platforms such as Excel and SharePoint are standard by now and are still the prevailing methods of deployment in the field. Excel-based deployment allows further manipulation of the results, similar to OLAP-based deployment; and
- The creation of a variety of custom applications using APIs.

The selected deployment model will often be based on the level of sophistication of the IT systems employed by a given organization.

5.8 CRISP-DM: Summary

Chapter 5 described the application of the CRISP-DM methodology to the creation of the data mining models that were based on the answers received from the questionnaires. While, in the majority of the models, the six competence areas were used as a grouping of input variables, such a categorization of replies is not mandatory in order to successfully use DM methods in the generation of knowledge about organizational OR. (The clustering model is an example of the model that creates groupings of questions based on the algorithm's interpretation of the data. It would be possible to use algorithm-generated groupings, in place of the six competence areas, in the construction of other DM models.)

Thereafter, the chapter addressed each individual section of the CRISP-DM model as it was applied to this research. Or, in other words, this research was presented in the CRISP-DM context.

The 'Business Understanding' section addressed the needs of a business, such as identifying business goals, assessing its current situation and forming its DM goals, and considered the needs of this research. As such, a number of key objectives were identified that guided the development of the research throughout the remaining sections of Chapter 5. (From the business perspective, it was very pleasing to illustrate the practical nature of this research with respect to DM, KM and OR. Methods for obtaining a numerical 'OR Score' and classification organizations into 'resilient/not resilient' were illustrated.)

The 'Data Understanding' phase addressed the need to understand the data, identify problems with the data and missing data in order to anticipate problems in the modeling phase. The most common data issues, along with the correction mechanisms, were briefly discussed. General statistics of the data collected via the questionnaires were also presented and discussed.

The section concerning 'Data Preparation' discussed various techniques for addressing the needs of data cleaning and data transformation so that the data used in the mining algorithms produces meaningful results. The discussion of data transformation was further expanded due to the needs of this research, primarily in the area of the variable data type: there was a need to transform numerical variables into categorical ones, reduce the dimensionality of the data and to discretize continuous variables.

The 'Modeling' section of this thesis, instead of discussing the models created in this research, considered general concepts that apply to all DM models. Because of the logical structure of this thesis, which combines the presentation of models with findings and discussion, the models themselves are presented in Chapter 6.

The 'Evaluation' section expanded on the discussion of the evaluation of the predictive abilities of the DM models when the topic of the first Naïve Bayes model was introduced. Clearly, the insufficient amount of input data did not allow for the carrying out of a detailed and meaningful evaluation; however, evaluation techniques were presented and discussed. The concepts of lift charts, scatter plot graphs, classification matrices and cross-validation were discussed in relation to some of the models created in this research. The critical concepts of false positives and false negatives, among other outcomes, were explained.

The last section of the chapter prior to the summary section, the 'Model Deployment' section, was added to ensure the completeness of the presentation of the CRISP-DM framework, as the deployment of the model developed was not a part of this work. The section briefly explained methods for sharing the knowledge generated by the DM models.

In the next chapter, the five DM models used in this research are presented and discussed (two Naïve Bayes models are discussed in Section 6.2). The next chapter begins by introducing some of the concepts used in the later sections; hence, it is somewhat lengthier than introductions to other chapters. Each of the four sections of Chapter 6 that discuss the DM models follows the same format. First, there is an introduction to the section and to the algorithm used in it. Then, issues specific to the particular algorithm are discussed. The construction of the model takes place next, and each section finishes with the findings and then a discussion in the context of research questions # 3, #4 and #5. That is, each section is self-contained and is composed of a model presentation, findings and discussion. This presentation structure has been selected because the applied nature of the research does not fit well with the traditional basic research format that clearly separates findings, discussion and conclusions. When used in this work, the traditional layout resulted in fragmented sections that were, consequently, hard to follow.

CHAPTER 6: DATA MINING

6.1 Introduction

This chapter presents the DM-related findings of this research, addressing research questions #3, #4 and #5. The models are presented in terms of the DM algorithm they use to generate results and include two Naïve Bayes models, presented in Section 6.2; a clustering model, presented in Section 6.3; a neural network model, discussed in Section 6.4; and the decision trees model, which is presented in Section 6.5. Justification for the selection of the algorithms investigated in this chapter is provided in Section 5.5.2.

Each chapter discussing DM model has a similar format: It will first discuss data/process requirements specific to a given model and will then move on to the model presentation, seeking to answer research questions #3 and #5. (RQ #3 and RQ #5 symbols in the text designate the areas that affect the corresponding research question).

The findings section (6.7) discusses and summarizes the findings of this research in the context of research questions #3, #4 and #5.

6.2 Data Mining: Naïve Bayes (NB)

As the name implies, the NB algorithm derives its name from the Reverend Thomas Bayes, the English mathematician and Presbyterian minister, who is viewed by many in the industry as the 'father of modern machine learning', thanks to, among other contributions, his arrival at the Bayes' Theorem in the 1740s.

As stated by MacLennan et al. (2009, pg. 216), the NB 'algorithm learns the evidence by counting the correlations between the variable you are interested in and all other variables.' However, despite its simplistic approach, the NB algorithm can achieve impressive results, rivaling more sophisticated classifier algorithms (Witten et al., 2011, pg. 99).

It is a common industry practice to use the NB algorithm first in order to learn more about the data to be used for analysis; however, it is important that the limitations of the NB algorithm are acknowledged and properly addressed. Witten et al. (2011, pg. 99) strongly encourage the use of the NB as the first DM models, as do MacLennan et al. (2009, pg. 217).

One of the key drawbacks of the NB algorithm is the fact that it considers each input attribute independently of all others (Abbott, 2014, pg. 270; MacLennan et al., 2009, pg. 217; Witten et al., 2011, pg. 90; Kuhn & Johnson, 2016, pg. 356). The other two weaknesses of the NB algorithm are that it requires categorical inputs and that it does not discover interactions in the data (Abbott, 2014, pg. 270). The algorithm received the 'naïve' name primarily because of the 'independence limitation' mentioned above.

For the purpose of this research, the following two NB models were constructed:

- NB_Model1 – In this model, six competence areas (discussed in Section 4.6.2) and the dependent variable 'OR Discretized' were used both as input and as output. This technique is common in the industry for learning about the input data. MacLennan et al. (2009, pg. 217) state that '[a] good way to start mining data is to create a Naïve Bayes model and check both input and predictable on all non-key columns. The resultant model provides you with a better understanding of your data and helps you build better subsequent models.' The purpose of building this model was to learn about the interactions of each of the competence areas on another and on the dependent variable.
- NB_Model2 – In this model, six competence areas were also used, but, this time, the competence areas functioned solely as independent variables (they were marked as an input-only type of variable). The 'OR Discretized' was the only dependent variable. The purpose of this model was to analyze, considering each competence area independently, what makes an organization resilient – the key input variables and the composition of the resultant 'levels of OR'.

  Note that the 'levels of OR' are the numerical ranges of the dependent variable, sometimes also referred to as the 'OR Score' or 'OR Node'.

Prior to the creation of the first NB model, a few preliminaries need discussion.

6.2.1 Naïve Bayes Preliminaries

One of the practical considerations when employing the NB algorithm, as stated by Abbott (2014, pg. 269), is the need for categorical inputs. Because the use of the numeric fields in the building of the NB model did not result in the construction of a model (the constructed model contained a single, 'OR' node instead of one node per each input parameter), the categorical values were assigned into the fields that could hold categorical values. These categorical data type fields were CreateStr, ExploitStr, DecideStr, LearnStr, ConnectStr, LinkStr, PerformanceStr and ORStr, with the first six fields corresponding to the six competence areas, 'PerformanceStr' corresponding to the performance aspect of an organization and 'ORStr' corresponding to the output variable.

According to Larose and Larose (2015, pg. 41), there are four common methods for the conversion of numerical variables into categorical types:

- Equal width binning – dividing the numerical predictor into a pre-selected number of equal width categories;
- Equal frequency binning – dividing the numerical predictor into categories based on the equal number of records in each category;
- By clustering – using a clustering algorithm to automatically determine optimal partitioning; and
- By predictive value – partitioning the numerical predictor based on the effect each partition has on the value of the target variable.

For the purpose of this research, an approach similar to 'equal width binning' was used as a method of converting between numerical and categorical data types. The equal width binning method, with five categories, 'A' to 'E', appeared to be a good match to the scale of response (which consisted of five responses, from 'strongly disagree' to 'strongly agree'). It can be also expected that any firm responding with a minimal 'number of points' (the case when they respond 'strongly disagree' to every question) would be classified in the lowest band, 'E.' Similarly, an organization responding 'disagree' to every question would be expected to be classified in the penultimate band, 'D'. The same argument applies to classification into bands 'C,' 'B' and 'A'.

Equal frequency binning' was not selected because the creation of records based on an equal number of elements in each collection seemed to be unnatural, and it also assumes that each category is equally likely to occur.

As stated by Larose and Larose (2015, pg. 41), 'by clustering' and 'by predictive value' are the preferred methods for binning. However, the very small data sample and the lack of emphasis placed by this thesis on the actual results returned by the DM models make the methods suggested by Larose and Larose difficult to justify in this research. Also, based on the personal experience of the author of this research, binning 'by clustering' and 'by predictive value' are not necessarily the first options used in the industry, as simpler solutions appear to have precedence.

Using the 'equal width binning' method and the 'A' to 'E' scale to categorize numerical variables into 'categorical type' that were based on the values of the fields containing a ratio of points collected over points possible to collect within a specific section. That is, the values of fields CreateRatio, ExploitRatio, DecideRatio, LearnRatio, ConnectRatio, LinkRatio, PerformanceRation and ORRatio were transformed into the values stored in the corresponding, one-to-one, fields: CreateStr, ExploitStr, DecideStr, LearnStr, ConnectStr, LinkStr, PerformanceStr and ORStr. The following formula was used in the conversion process:

Value of 'Ratio' field:	Assigned categorical value:
0.00 – 0.2	E
0.21 – 0.4	D
0.41 – 0.6	C
0.61 – 0.8	B
0.81 – 1.0	A

Table 6.2.2.1: Initial binning attempt

Upon the inspection of the outcomes of the conversion using the rules specified above, it was clear that there were no entries assigned to the 'E' category and only one 'D' entry in the LinkStr field for Organization IP = 30, making the 'E' and 'D' categorical values of little value.

To make the outcomes of the assignment of the numeric values into categorical values more uniform, new binning rules were developed and applied:

Value of 'Ratio' field:	**Assigned categorical value:**
Less than 0.60	F
0.60 - 0.69	D
0.70 – 0.79	C
0.80 – 0.89	B
0.90 – 1.0	A

Table 6.2.2.2: Intermediate binning results

Finally, a new column (called 'ORIntDiscretized') was added to the table holding the replies to the questionnaire 'tbl_DM_KM_OR_RGU'. This newly added column, of the type tinyint, was an integer representation of the value contained in the 'ORRatio' field (applying the so-called 'numerosity reduction' introduced in Chapter 5.4). The value assignment in this newly added column (functioning in the NB model as the dependent variable) used the rounding method, according to the following rules:

Value of 'Ratio' field:	**Assigned ORIntDiscretized:**
Less than 0.55	5
0.55 – 0.65	6
0.65 – 0.74	7
0.75 – 0.84	8
>= 0.85	9

Table 6.2.2.3: Final values in new column ORIntDiscretized

The assignment rules were slightly different from the rules for the categorical assignment. Setting the 'highest' bracket at >= 0.85, instead of >= 0.90, provided an opportunity to assign more entries into the 'top bracket,' in order to facilitate learning for the DM algorithm.

The decision to add the 'ORIntDiscretized' column was influenced by the output of the NB algorithm. Initially, when the column 'ORStr' holding categorical values was used as the dependent variable, the resultant NB model contained only the single dependent variable node: 'OR Str' (illustrated in Appendix VIII, Fig. A8.6.) With the values of the column being between 5 and 9, the resultant

NB model could have been specified as ‘discrete,’ resulting in only 5 possible outcomes: ‘5’ through ‘9’. Otherwise, there were two, not very desirable, representations of the outcome:

1. If the result were to be specified to be of ‘discretized’ type, the resultant groups discretized by the DM algorithm would not match the values of the output variables well; or
2. If the result were to be left to be of the original, the integer type, there would be one group per possible output value (roughly 40 groups, from values 50 through 90, corresponding to each value of the output variable). Fig. A8.7, in Appendix VIII, provides a pictorial representation of setting the content type for a model.

Since NB is a classifier-type of algorithm, it follows the general approach to classification, which is the two-step process. As stated by Han et al. (2012, pg. 328) ‘[d]ata classification is a two-step process, consisting of a learning step (where the classification model is constructed) and a classification step (where the model is used to predict class labels for given data.’ This two-step process is illustrated in the next two sections, with the additional information included in Appendix VIII.

6.2.2 Naïve Bayes Models

NB_Model1 Model

NB_Model1, the model that uses all six competence areas discussed in Section 4.6.2 and the dependent variable (OR: ORIntDiscretized) as both an input and output, is discussed next. As stated earlier, such practice (MacLennan et al., 2009, pg. 217) is used to learn about the relationships between attributes. In the case of this research, the primary purpose is to learn about the relationships between each competence area (group of KM processes) and between the dependent variable (OR) and each competence area in order to be able to address research questions #3 and #5. The DM model was built using the Data Mining Wizard (part of SQL Server 2012), accepting the wizard’s default values, such as a hold-out of 30% of the data for model testing, in the building process. All of the columns used in the model were designated as ‘Discrete’ when asked by the wizard to specify the column content’s data type.

The NB_Model1 (as well as the NB_Model2, discussed later) used is presented in Appendix XI and in Fig. A8.9 in Appendix VIII.

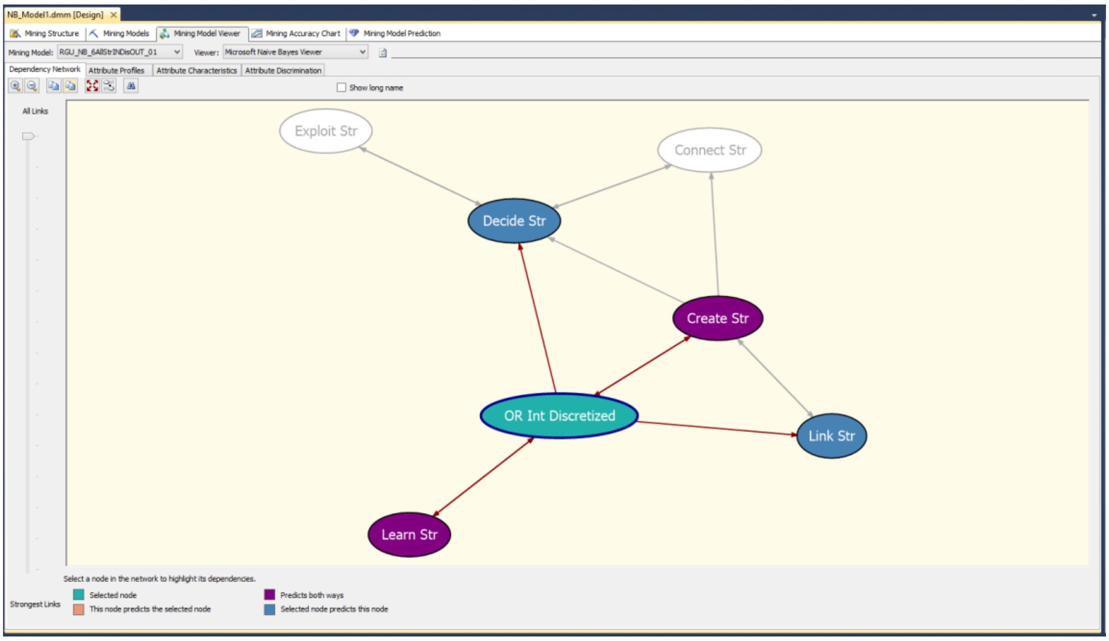
Fig. A.8.10 in Appendix VIII illustrates the mining model constructed. Each column of the structure, with the exception of the key column 'IP,' has been set to be of 'Predict' usage, meaning that the column is to be used, as an input, in predicting the other column that was set to the 'Predict' usage. That is, if the column 'Connect Str' is set to the 'Predict' usage and the column 'Create Str' is also set to the 'Predict' usage, then the column 'Connect Str' will be used as an input in predicting 'Create Str.' (Other choices for the selection of data column usage, per Larson [2012, pg. 625] include Key, a unique identifier for a table; Input, an input data column used by the DM algorithm to make a prediction; Predict Only, a column for which the value is predicted by the DM algorithm; and Ignore, a column not to be used by the DM algorithm).

Each data mining algorithm that is a part of the MS SQL Server 2012 suite has its own associated viewers, allowing for the inspection of the algorithm's outcome. In addition, each DM algorithm has a set of parameters controlling its operation/output. When necessary and where different from the system default value, the value of the parameter is discussed. The following key points about the viewer, when applied to the NB_Model1, need to be made. (The outcome of the NB_Model1 can be seen below, in Fig. 6.2.3.1).

It is worth noting that the NB algorithm's parameter MINIMUM_DEPENDENCY_PROBABILITY, specifying the dependency probability '0 to 1' between input and output variables, was set to 0.51 (Fig. A8.11 in Appendix VIII). Unless otherwise stated, this parameter value is assumed in the discussion in this section. The reason for setting the value of the parameter to 0.51 is that 50% (0.50) represents a 50-50 chance for the existence of dependencies between the variable, and the model should consider, at least, slightly better probability than 50%. Cleary, the higher the value of the parameter, the more profound the relationships are between the variables displayed in the viewer (RQ #3); however, with the limited data, setting the value too high may produce no model at all. (Of the remaining variables, MAXIMUM_INPUT_ATTRIBUTES specifies the maximum number of input attributes that the algorithm can handle before invoking further optimization;

MAXIMUM_OUTPUT_ATTRIBUTES specifies the maximum number of output parameters that the algorithm can handle before invoking further optimization; and MAXIMUM_STATES specifies the maximum number of attributes that the algorithm supports – the values of these parameters have been left with the default values, which fully support the input/output data used in this research. No new, 'custom' parameters were introduced to the model.)

The dependency network displayed in the 'NB viewer' for NB_Model1 is shown in Fig. 6.2.3.1. This model enables the viewing of dependencies between all of the variables used in the model (RQ #3). The node with its name written inside it represents the variable, and the one- or two-directional arrows represent the relationship between the nodes. (For single directional arrows, the arrow is drawn from the attribute that is a predictor to the attribute it predicts.)



NB_Model1.dmm [Design]

Mining Model: RGU_NB_6AllStrINDisOUT_01

Viewer: Microsoft Naive Bayes Viewer

□ Show long name

Select a node in the network to highlight its dependencies.

Fig. 6.2.3.1: Dependency network @ MINIMUM_DEPENDENCY_PROBABILITY = 0.51

As shown in Fig. 6.2.3.1, the relationships between the key variables (as all variables are used to predict all other variables) indicate the following:

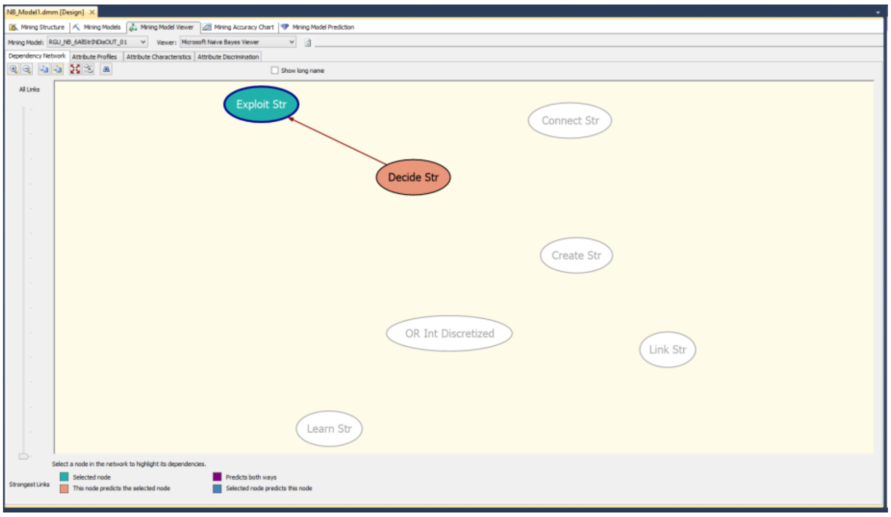
- Create and Learn competences predict OR ('OR Int Discretized') and vice versa;
- OR ('OR Int Discretized') predicts Decide and Link competence areas; and

- Exploit and Connect competences predict OR ('OR Int Discretized') and vice versa.

Using the slider located in the left portion of the screen allows the limiting of the display of links to the desired viewing 'strength': from 'All Links' to the 'Strongest Link'. Fig. 6.2.3.2 illustrates the model with the slider in the 'Strongest Link' position, which shows that the Decide competence 'predicting' the Exploit competence to be the strongest.

Because of the key interest of this research in possible ways of determining the impact of all competence areas on OR (RQ #3), the slider measuring the strength of the links has been moved to the position which shows the 'dependency link' to/from the 'OR Int Discretized' node. This scenario is illustrated in Fig. 6.2.3.3.

As shown in Fig. 6.2.3.3, the model with the limited data indicates the dependency network that exists between the 'OR' node and the nodes 'predicted' by it, the Decide, Link and Learn competences (such an interpretation makes sense, as all parameters have been set to be input/output parameters). Also worth pointing out is the very strong dependency between the Connect and Decide competences, with the Connect competence predicting the Decide competence.

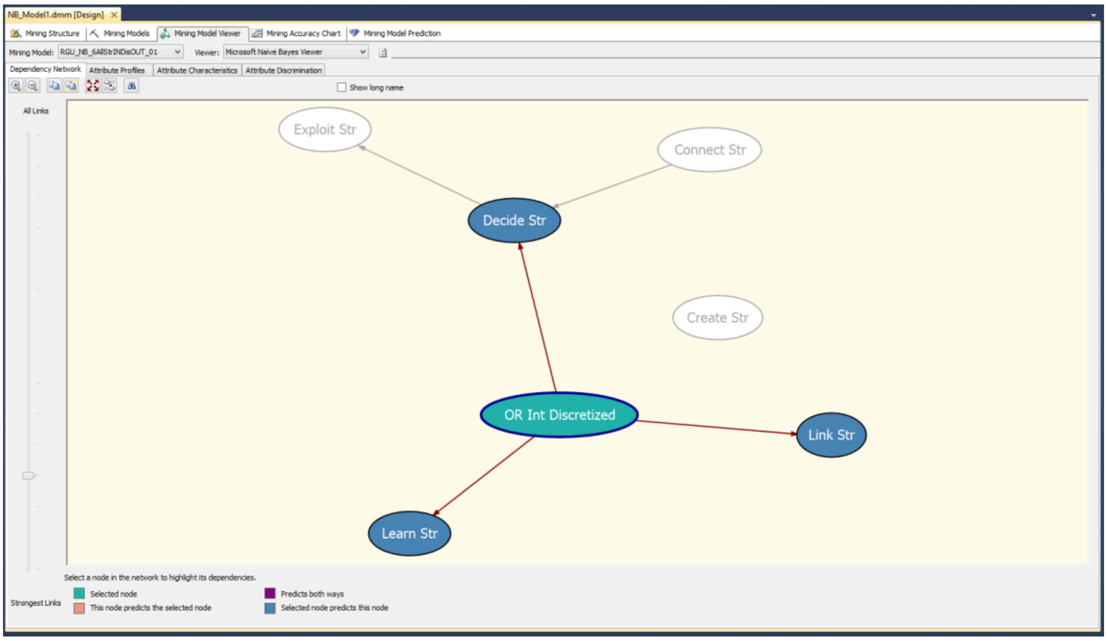


Mining Model: RGU_NB_6AllStrINDisOUT_01

Viewer: Microsoft Naive Bayes Viewer

Select a node in the network to highlight its dependencies.

Fig. 6.2.3.2: Dependency network displaying 'the strongest' link



Select a node in the network to highlight its dependencies.

🟩 Selected node
🟥 This node predicts the selected node
🟪 Predicts both ways
🟦 Selected node predicts this node

Fig. 6.2.3.3: Dependency network with the ‘strongest’ links leading to/from OR node

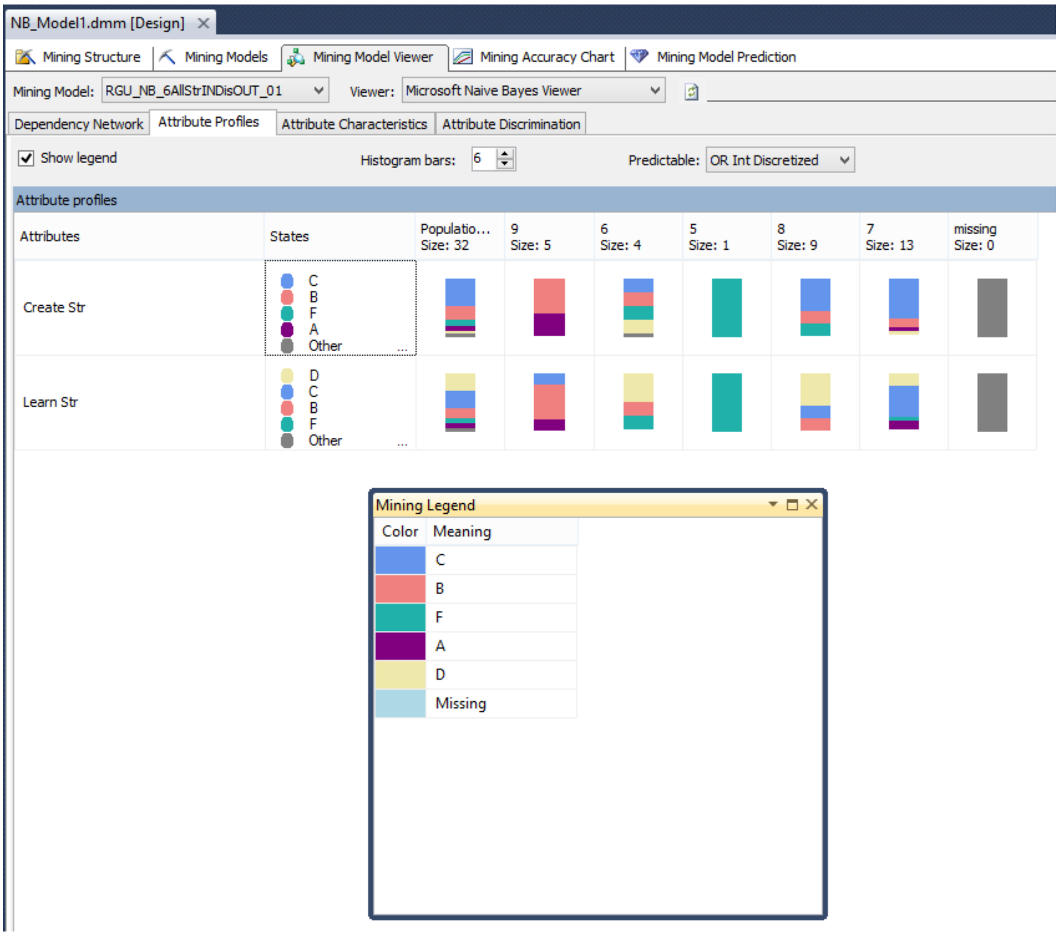
Corresponding to the dependency network are the attribute profiles that are present on one of the tabs of the NB viewer. This functionality of the viewer enables an investigation into which values and attributes contribute to a specific outcome. As described by Larson (2012, pg. 647), the Attribute Profiles option presents a view of how each input attribute corresponds to each output attribute, displaying one attribute at a time.

Fig. 6.2.3.4 shows the attribute profiles for predictable ‘OR Int Discretized’. Per Fig. 6.2.3.1, the two strongest predicting ‘OR Int Discretized’ nodes are the ‘Create Str’ and the ‘Learn Str’, corresponding to the Create and Learn competences. The two predicting the ‘OR Int Discretized’ variables ‘Createa Str’ and ‘Learn Str’ are listed in Fig. 6.2.3.4 in row-wise fashion (RQ #3). The values of the predicted variable are listed in column-wise fashion, divided into five groups (from 5 to 9), one group per single value (enforced by the selection of the data type of the ‘OR Int Discretized’ to be of ‘Discrete’ type). Also listed are the characteristics of the input set (which amounted to 32 records after holding some records for model testing) as well as the information about any ‘Missing’ values; there are no missing entries in any of the models, per the discussion in Sections 5.3 and 5.4.

Note that, from the professional experience of the author of this work and various instructional materials available on the Internet and in print, when viewing the attributes profiles it is important to establish the appropriate context: to view the attribute's characteristics across all outcome values and to look at each outcome value across all attributes. Two other points about attribute profiles made by MacLennan et al. (2009, pg. 227) are as follows: 'First, an attribute characteristic does not imply predictive power. Second, inputs that fall below the minimum node score in the algorithm parameters are not displayed'. In the case of the second point, with the number of histogram bars set to 6, when the output value of '9' and the input 'Create Str' is inspected, it can be seen that the bar shows only two values (two colors, respectively): the larger one for input state = 'B' and the slightly smaller one for state = 'A'. All other states failed to reach the 0.51 dependency probability value.

As seen in Fig. 6.2.3.4, an inspection of the values of the output attribute '9' (the most desired) holding five members points that 60% of 'Create Str' variable contains converted to categorical value answer 'B' and 40% of answer 'A'. (The percentage values are displayed on the screen upon moving the mouse cursor over the histogram). The least desired output value stored in the column, labeled '5,' containing a single member for both input variables, leading to a value of 'F'.

The 'Attribute Characteristics' tab of the NB viewer allows inspection of the characteristics of each attribute predicting it, still subject to restriction by the values of the algorithm's parameters (RQ #3, RQ #5). That is, using as a reference Fig. 6.2.3.1, where 'Learn Str' and 'Create Str' predict 'OR Int Discretized' node as being the strongest, the 'Attribute Characteristics' viewer allows for investigating the probability of the value of the attribute to contribute to the specific value of that variable. For example, selecting the attribute to be 'OR Int Discretized' and the inspection value = 5, it can be seen (Fig. 6.2.3.5) that the two variables predicting it, the 'Learn Str' and 'Create Str', have to each assume the value of 'F' in order for it to result in an ending value of 5 (or in 100%, seen when the mouse is moved over the dark blue line).



NB_Model1.dmm [Design]

Mining Model: RGU_NB_6AllStrINDisOUT_01 Viewer: Microsoft Naive Bayes Viewer

☑ Show legend Histogram bars: 6 Predictable: OR Int Discretized

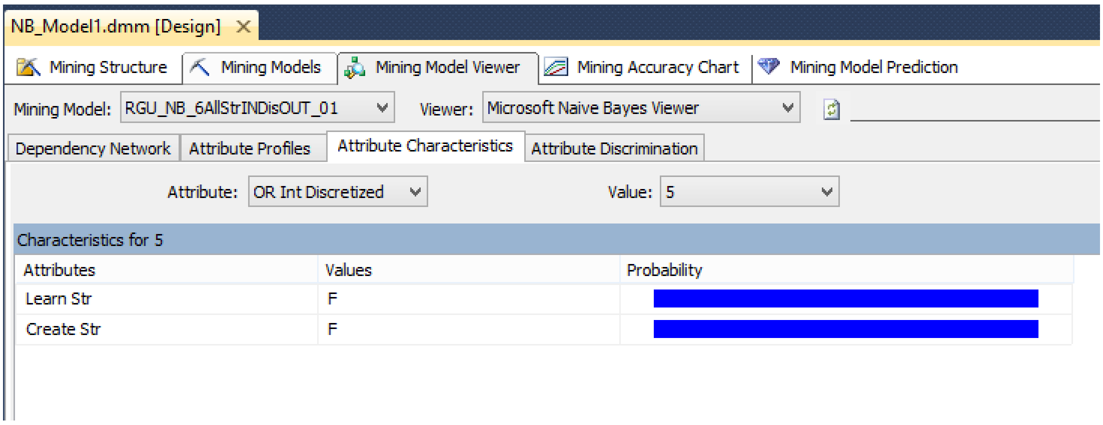
Attribute profiles

Attributes	States	Populatio... Size: 32	9 Size: 5	6 Size: 4	5 Size: 1	8 Size: 9	7 Size: 13	missing Size: 0
Create Str	C B F A Other							
Learn Str	D C B F Other							

Mining Legend

Color	Meaning
🟦	C
🟥	B
🟩	F
🟪	A
🟨	D
🟦	Missing

Fig. 6.2.3.4: Attribute profile for predicable 'OR Int Discretized'



NB_Model1.dmm [Design]

Mining Model: RGU_NB_6AllStrINDisOUT_01 Viewer: Microsoft Naive Bayes Viewer

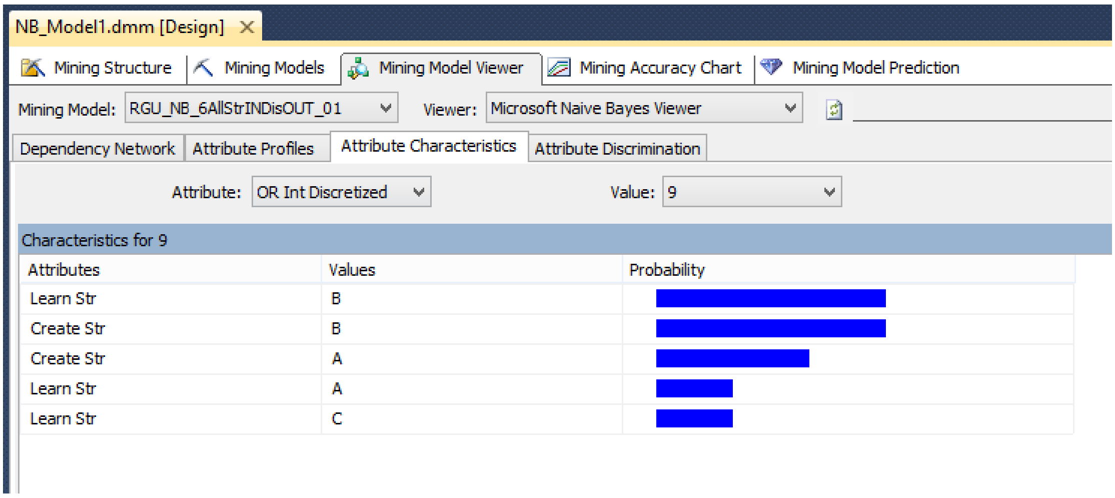
Attribute: OR Int Discretized Value: 5

Characteristics for 5

Attributes	Values	Probability
Learn Str	F	
Create Str	F	

Fig. 6.2.3.5: 'OR Int Discretized' variable with value = 5

Similarly, in Fig. 6.2.3.6, the values for both input variables are shown, along with the probabilities of achieving an output value of '9'. (Not easily visible from the graph, but well described upon placing the mouse's cursor on the appropriate area, are the probabilities associated with the five listed attributes, which are, from top to bottom, 60%, 60%,40%, 20%, 20% – the lowest probability of 20% is given to the 'Learn Str' variable for holding the values of 'A' and 'C'.



Mining Model: RGU_NB_6AllStrINDisOUT_01

Viewer: Microsoft Naive Bayes Viewer

Attribute: OR Int Discretized

Value: 9

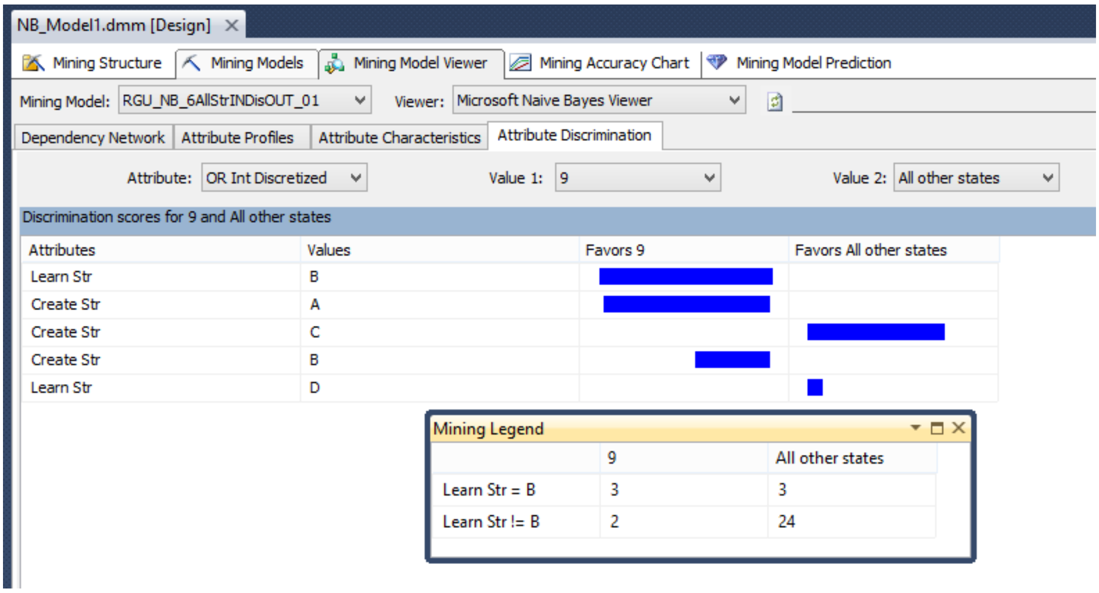
Characteristics for 9

Attributes	Values	Probability
Learn Str	B	▇▇▇▇▇▇
Create Str	B	▇▇▇▇▇▇
Create Str	A	▇▇▇▇
Learn Str	A	▇▇
Learn Str	C	▇▇

Fig. 6.2.3.6: 'OR Int Discretized' variable with value = 9

The last tab, 'Attribute Discrimination,' provides the answers to what is perhaps the most interesting question: What is the difference between A and B, or, in the case of this research, what is the difference between 'OR Int Discretized' with a value of '5' and those equal to '9' (RQ # 3)? With this viewer, one needs to choose the attribute of interest and the state of interest for the selected attribute. According to MacLennan et al. (2009, pg. 228), '[y]ou can determine the unique characteristics of a group by comparing one state to all other states. This will give you a view of what separates the particular group from the rest of the crowd.'

Fig. 6.2.3.7 presents the discrimination of the 'OR Int Discretized' variable by comparing state = 9 with the 'all other states'. Based on the graphs shown, a 'Learn Str' attribute favors value B, and 'Create Str' attribute favors the value A in order to be a part of the output 'OR Int Discretized' = 9. On the other hand, the value of C of the 'Create Str' attribute favors all states other than 'OR Int Discretized' = 9, so it will not be easily found in that output group.



Mining Model: RGU_NB_6AllStrINDisOUT_01 Viewer: Microsoft Naive Bayes Viewer

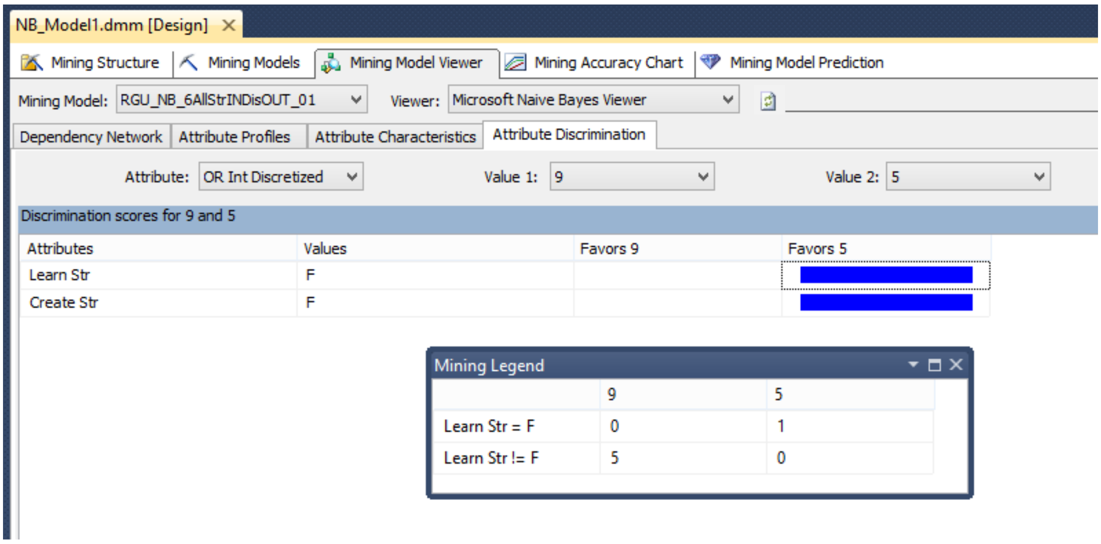
Attribute: OR Int Discretized Value 1: 9 Value 2: All other states

Discrimination scores for 9 and All other states

Attributes	Values	Favors 9	Favors All other states
Learn Str	B	■	
Create Str	A	■	
Create Str	C		■
Create Str	B	■	
Learn Str	D		■

	9	All other states
Learn Str = B	3	3
Learn Str != B	2	24

Fig. 6.2.3.7: ‘OR Int Discretized’ attribute discrimination: value of ‘9’ vs. ‘all other’



Mining Model: RGU_NB_6AllStrINDisOUT_01 Viewer: Microsoft Naive Bayes Viewer

Attribute: OR Int Discretized Value 1: 9 Value 2: 5

Discrimination scores for 9 and 5

Attributes	Values	Favors 9	Favors 5
Learn Str	F		■
Create Str	F		■

	9	5
Learn Str = F	0	1
Learn Str != F	5	0

Fig 6.2.3.8: ‘OR Int Discretized’ attribute discrimination: value of ‘9’ vs. value of ‘5’

Appendix VIII, Fig A8.12 and Fig A8.13 provide additional examples of ‘Attribute Discrimination’.

When inspecting the attribute discrimination between two attributes, it is critical to ensure there is a support level in place (MacLennan et al, 2009, pg. 228); that is, that there is a sufficient number of cases supporting the discrimination (RQ #5). The number of cases can be seen in the mining legend box displayed in figures 6.2.3.7 and 6.2.3.8. (Clearly, in both cases pictured, the support is significantly too low for any prudent prediction to be made by any model included in this work.) Finally, care is needed when interpreting the results of the attribute discrimination. It is not that the belonging of the attributes to one or the other group is implied; rather, it is implied that these factors favor one group over another (RQ #5). Larson (2012, pg. 648) refers to 'Attribute Discrimination' in the following fashion: 'This diagram lets us determine what attribute values most differentiate nodes favoring our desired predicable state from those disfavoring our predictable state.'

Note that any one of the seven attributes (the six competence area attributes and the 'OR Int Discretized' attribute) can be investigated in the fashion described above (RQ #3), bearing in mind the concept driving the analysis of predicting variable(s) and any custom values in the algorithm's parameters (RQ #5).

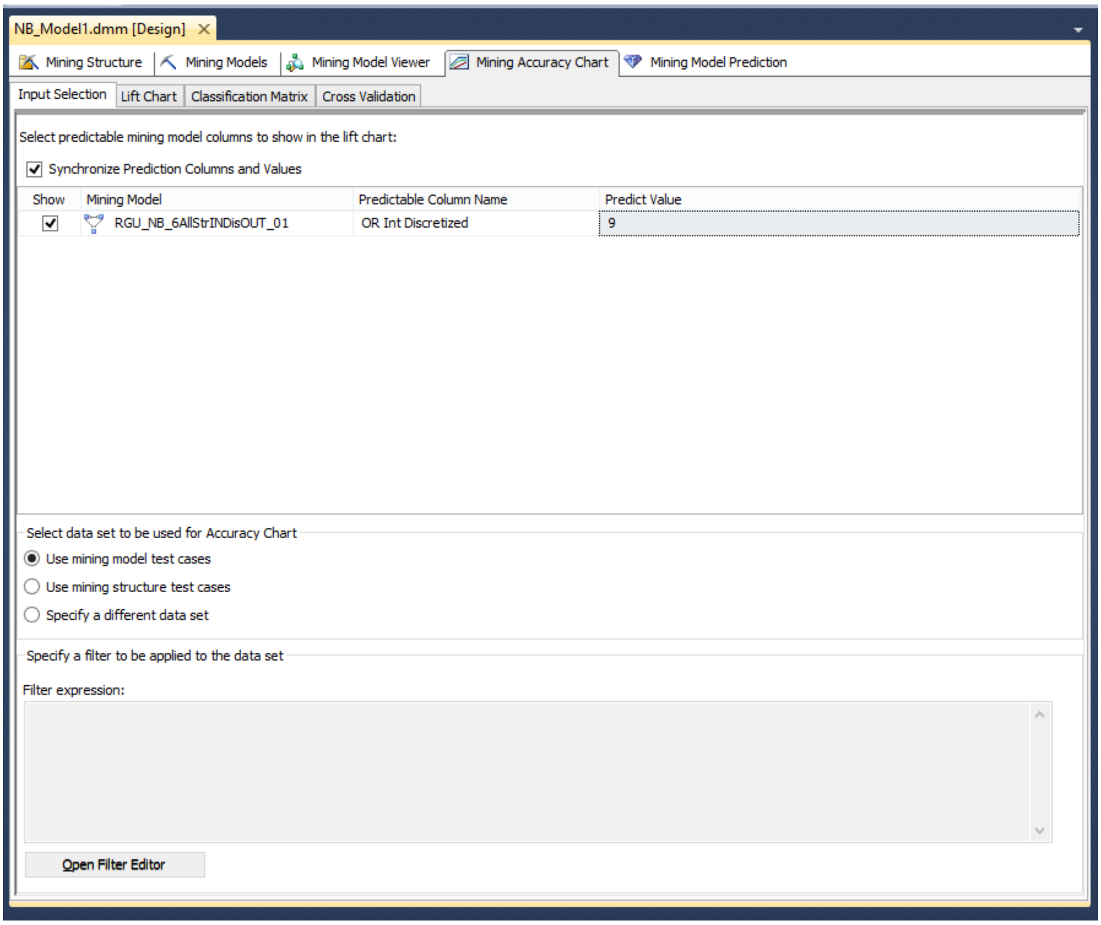
The mining accuracy chart area, which evaluates a predictive model that is not based on a time series or association rules algorithm, is composed of four sections:

- Input section;
- Lift chart;
- Classification matrix; and
- Cross validation

The mining accuracy chart available in MS SQL Server 2012 compares the predictive capability of the predictive model (in this case the Naïve Bayes model) to both an ideal model achievable from the input data and an average model that achieves 50% accuracy with 50% of the data.

In order for the predictive model to have its performance evaluated, some part of the input data needs to be held for testing. (In the case of the model-building conducted for the needs of this research, the default 30% of the data was set aside for testing. The percentage of the data to be withheld is a part of the model-building wizard illustrated in Appendix VIII.) As stated by Abbott (2014, pg. 123), using the same data for testing and training may indicate that the model performs better than it actually does due to the condition called overfitting (a concept that was discussed in Section 5.6.1).

As presented in Fig. 6.2.3.9, the input section allows for selection of the prediction value to be tested. (There are three choices for the data set to be used in the accuracy chart. The first two options, 'Use mining model test cases' and 'Use mining structure test cases,' are equivalent to each other if there is no filter used with the second option. The third option, 'Specify a different data set,' allows for the use of a data set external to the model.) The 'OR Int Discretized' attribute has been selected as the 'Predictable Column Name' and the value of '9' selected as the value to be predicted.



Select predictable mining model columns to show in the lift chart:

☑ Synchronize Prediction Columns and Values

Show	Mining Model	Predictable Column Name	Predict Value
☑	RGU_NB_6AllStrINDisOUT_01	OR Int Discretized	9

Select data set to be used for Accuracy Chart

● Use mining model test cases

○ Use mining structure test cases

○ Specify a different data set

Specify a filter to be applied to the data set

Filter expression:

Fig. 6.2.3.9: The Mining Accuracy Chart sections

The lift chart, also referred to as the accuracy chart, (pictured in Fig. 6.2.3.10) illustrates the prediction capability of the model for predicting the value of '9' in the 'OR Int Discretized'. However, because the NB_Model1 has multiple variables selected as 'predictable,' the purpose of presenting the lift chart with relation to the NB_Model1 is to provide an introduction to the lift chart graph and not to analyze the outcome. A more detailed look at the lift chart and how to determine how well the model learns the patterns in the data is provided in Section 5.6 and Section 6.6.

For discrete types of target variable ('OR Int Discrete') standard, pictured in Fig. 6.2.3.10, lift chart displays are employed. This type of lift chart contains one line per evaluated DM model, in addition to the 45-degree line representing the random line (indicating that 50% of the target values can be predicted using 50% of the input data) and the ideal line (indicating in the picture that 9% of the input data would capture 100% of target values). Generally speaking, the ideal line indicates the model's upper (best performance) target line, and the random line indicates the lower (worst case scenario) meaningful outcome. Models at the random line or falling below it indicate that the DM model could not learn the patterns about the data from the training data set. 'Any improvement from the random guess (mining model's performance above the random line) is considered to be lift' (Microsoft Corp., 'Lift Chart [Analysis Services – Data Mining]').

The X-axis represents the percentage of the testing data set that was processed, and the Y-axis represents the percentage of the testing data that was used to make a correct prediction.

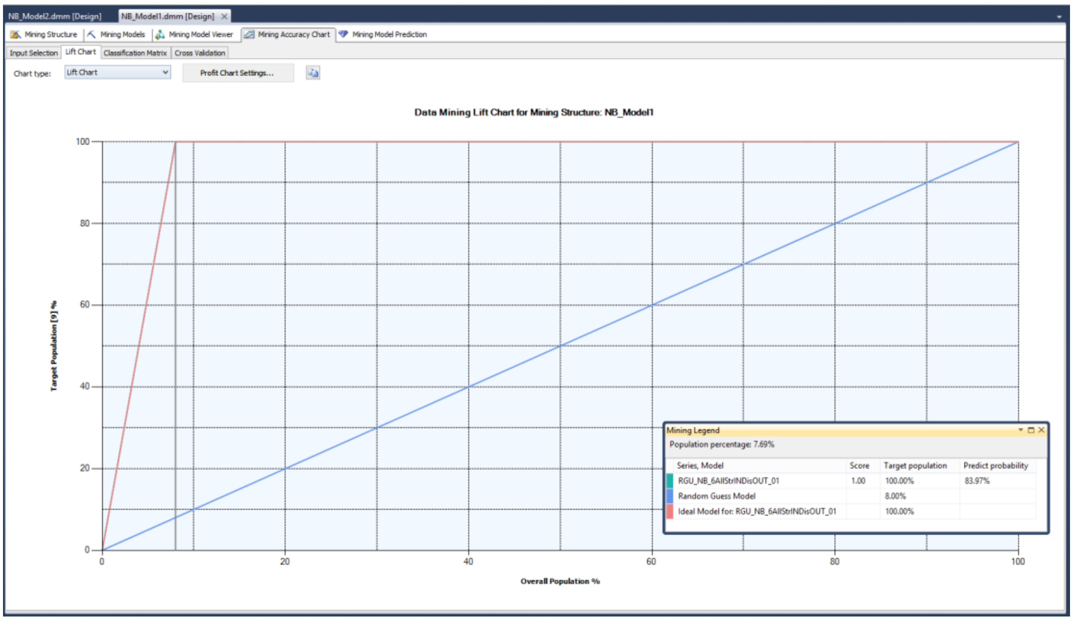
The vertical gray line serves the purpose of a marker, or reference point, when providing the DM results. It represents a certain overall population percentage against which the model performance is described. In the case of Fig. 6.2.3.10, the marker line has been set to touch the point where the ideal line predicts 100% of output values correctly (at 9% of the overall population, on the X-axis).

The mining legend provides the model's performance information at a certain overall population percentage (graphically, at the point where the vertical gray line is located). The legend provides statistics for each mining model considered.

According to the documentation about the lift chart presented on Microsoft's site (Microsoft Corp., 'Lift Chart [Analysis Services – Data Mining]),' the following are the meanings of the fields of the mining legend:

- The 'Series Model' column describes the elements evaluated by the lift chart in a model;
- The 'Score' column is used for comparison with other models;

- The 'Target Population' column indicates how much of the target population has been captured at the gray vertical line; and
- The 'Predict Probability' column displays the probability score that is needed for each prediction to capture the displayed target population.



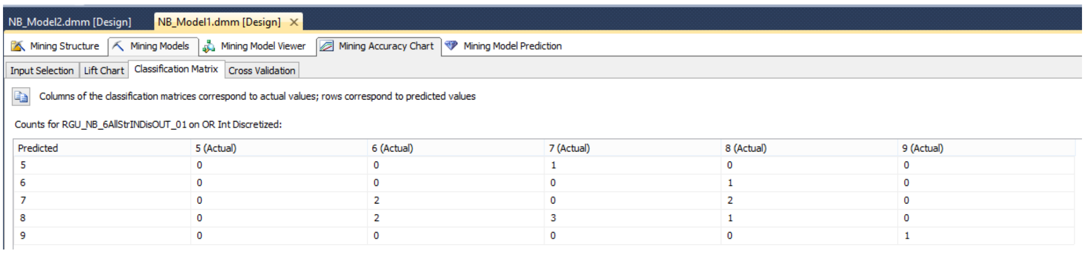
Mining Legend

Population percentage: 7.69%

	Series, Model	Score	Target population	Predict probability
🟩	RGU_NB_6AllStrINDisOUT_01	1.00	100.00%	83.97%
🟦	Random Guess Model		8.00%	
🟥	Ideal Model for: RGU_NB_6AllStrINDisOUT_01		100.00%	

Fig. 6.2.3.10: Lift chart for variable 'OR Int Discretized and value '9' of the mining model: NB_Model1

The classification matrix of the mining accuracy chart, presented below in Fig. 6.2.3.11, allows the details of the model's predictions to be viewed (including the mistakes made when predicting values). The columns represent the actual output values that were generated by the model; the rows illustrate what predictions were made for each one of the actual output values. Similarly to the lift chart, the classification matrix is discussed in greater detail in Section 6.6.



Columns of the classification matrices correspond to actual values; rows correspond to predicted values

Counts for RGU_NB_6AllStrINDisOUT_01 on OR Int Discretized:

Predicted	5 (Actual)	6 (Actual)	7 (Actual)	8 (Actual)	9 (Actual)
5	0	0	1	0	0
6	0	0	0	1	0
7	0	2	0	2	0
8	0	2	3	1	0
9	0	0	0	0	1

Fig. 6.2.3.11: Classification matrix for NB_Model1 model

The final tab of the mining accuracy chart, cross validation, is presented in Fig. 6.2.3.12. The cross validation is discussed in greater detail when each predictive DM model is presented. The cross validation technique examines the data and not the model, as was the case with the previous three tabs under the mining accuracy chart category and, as stated by Berthold and Hand (1999, pg. 56), '[c]ross-validation is a resampling technique that is often used for model selection and estimation of the prediction error of a classification – or regression function.'

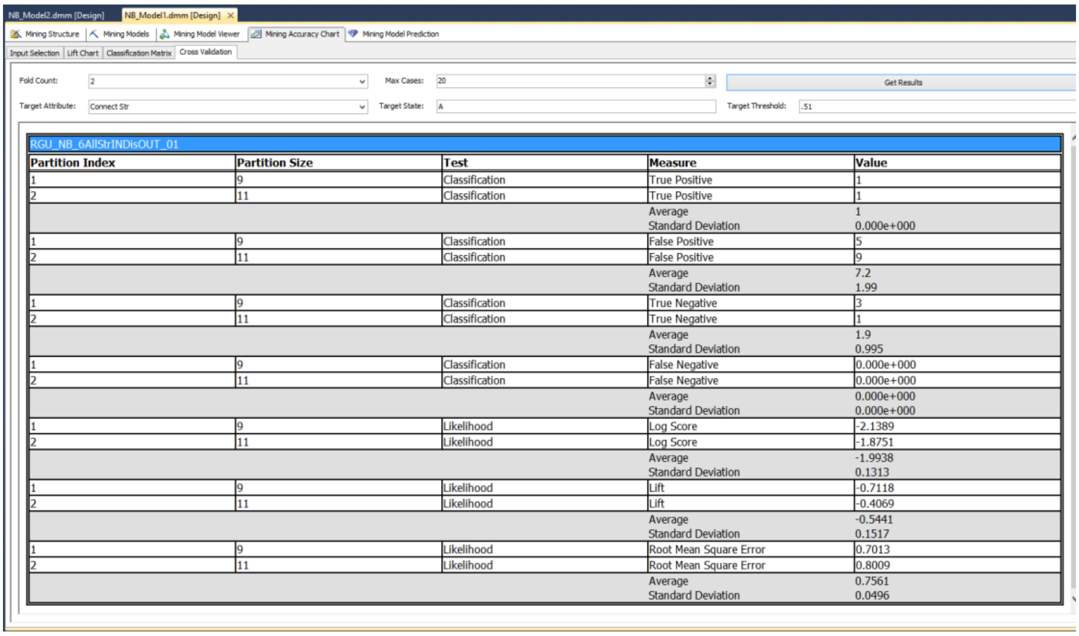
As illustrated by MacLennan et al. (2009, pg. 175), the cross-validation technique uses part (or all) of the model's training data. Then, it splits (or folds) the data into partitions that contain as many equivalent numbers of training cases as possible. Later, a mining model is built for each of the partitions, using the data from all of the other partitions, and the model is validated with the data of the current partition. According to MacLennan et al. (2009, pg. 175), the accuracy of the results returned by the validation needs to be investigated from two perspectives:

- The quality of the results – if the results are good, that provides a good indication of how good the training data is for the mining model. In the case where all of the partition models are of poor accuracy, the model trained with the data will also, most likely, be of low quality; and
- The results of the similar partitions – if the results vary greatly from model partition model to model partition, it indicates that there is insufficient data in the model. Differences suggest that partitions have significantly different data distributions.

The parameters for cross-validation are as follows (Larson, 2012, pg. 668):

- Fold count – the number of distinct sets to use;
- Max cases – the maximum number of cases to use for validation;
- Target attribute – the attribute to be predicted;
- Target state – the value of the 'Target Attribute' to predict; and
- Target threshold – the required probability that a prediction is correct (0.1..1.0)

It is a common industry practice to use the cross-validation method to determine which modeling technique will be the best for the task at hand, without the need to train the model for each algorithm (which can be both resource and time intensive).



Fold Count: 2
Max Cases: 20
Target Attribute: Connect Str
Target State: A
Target Threshold: .51

RGU_NB_6AllStrINDisOUT_01

Partition Index	Partition Size	Test	Measure	Value
1	9	Classification	True Positive	1
2	11	Classification	True Positive	1
			Average Standard Deviation	1 0.000e+000
1	9	Classification	False Positive	5
2	11	Classification	False Positive	9
			Average Standard Deviation	7.2 1.99
1	9	Classification	True Negative	3
2	11	Classification	True Negative	1
			Average Standard Deviation	1.9 0.995
1	9	Classification	False Negative	0.000e+000
2	11	Classification	False Negative	0.000e+000
			Average Standard Deviation	0.000e+000 0.000e+000
1	9	Likelihood	Log Score	-2.1389
2	11	Likelihood	Log Score	-1.8751
			Average Standard Deviation	-1.9938 0.1313
1	9	Likelihood	Lift	-0.7118
2	11	Likelihood	Lift	-0.4069
			Average Standard Deviation	-0.5441 0.1517
1	9	Likelihood	Root Mean Square Error	0.7013
2	11	Likelihood	Root Mean Square Error	0.8009
			Average Standard Deviation	0.7561 0.0496

Fig. 6.2.3.12: Cross-validation of NB_Model1

The mining model prediction builds on the steps described thus far and allows for making a prediction regarding OR. The mining model prediction interface is presented in Fig. 6.2.3.13, and the interface is further discussed when each model is discussed.

The mining model prediction area accepts two types of input data: a single set of input values (referred to as the ‘singleton query’) or the multiple input values (known as ‘prediction join’). For the purpose of this research, with the exception of the clustering algorithm, the predictions will be illustrated using the singleton query in attempt to find the ‘OR score’ (a single value between ‘5’ and ‘9,’ corresponding to the values found in the ‘OR Int Discretized’ variable, indicating the OR of an organization, with a higher value indicating better OR) value of a single organization rather than producing multiple predictions.

Among the practical outcomes of the ‘prediction phase,’ this work illustrates the possibility of using the DM models in order to obtain a single ‘OR score’ as well as to compare the resultant ‘OR scores’ received from various DM algorithms (RQ #3). The later part of this work consolidates all of the steps described above, leading to the model prediction phase along with their impact on the ‘OR score'. The discussion part of this thesis dwells further on the meaning of the 'OR score' in relation to KM.

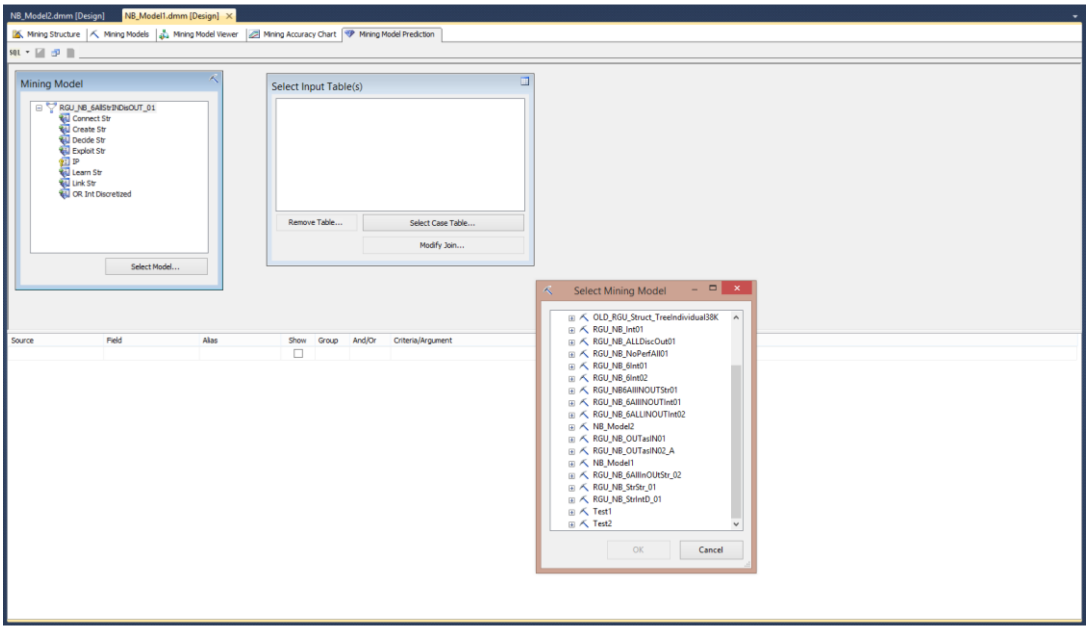


Fig. 6.2.3.13: Mining model prediction area

NB_Model2 Model

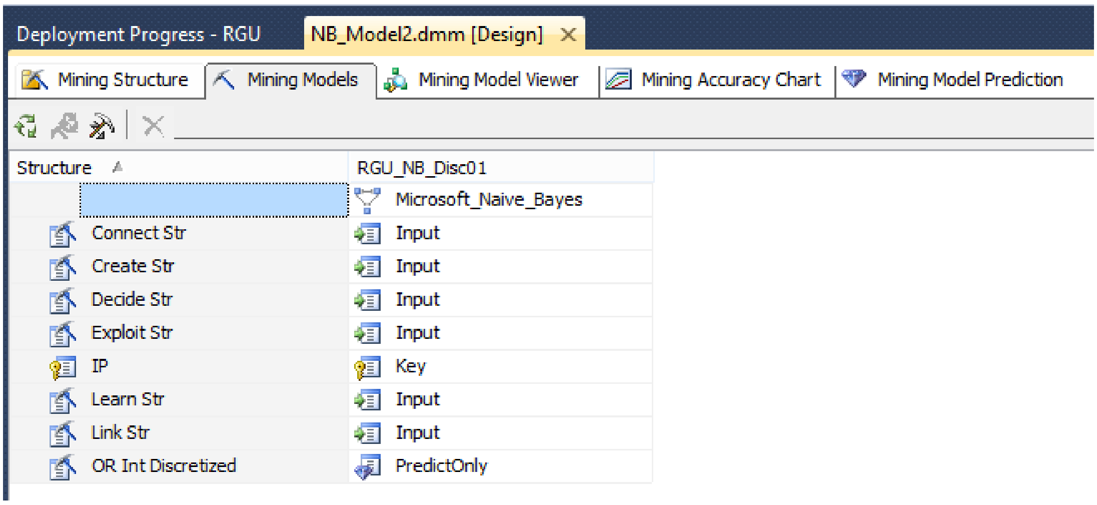
The second NB model, NB_Model2, uses the same data mining structure as the NB_Model1 model, consisting of the following attributes: ConnectStr, CreateStr, DecideStr, ExploitStr, LearnStr, LinkStr, IP and OR Int Discretized.

While evaluating the characteristics of the attributes based on the six competence areas and their impact on OR is part of the aims and objectives of this research, the NB_Model2 illustrates the practical, predictive capabilities of the model as well (RQ #3, RQ #5).

The NB_Model2, similarly to the prior model, was constructed using the Data Mining Wizard, accepting all of the default values during the construction, including the hold-out of 30% of data for testing. All of the columns used for the construction of the model were designated to be of the discrete type. However, the structure of the NB_Model2 differs significantly in the way the attributes (columns) are used, as in the NB_Model2 there is clear separation of the input and predict types of attributes. As shown in Fig. 6.2.3.14, all six competence areas constitute an input type of attribute. The 'OR Int Discretized' attribute is of the 'PredictOnly' type (meaning it has no impact on other attributes), and the

'IP' attribute is still the key of the model (uniquely identifying each data element or row in the table).

The purpose of this configuration of the model is to achieve a separation of the input versus output attributes for the purpose of prediction. The 'PredictOnly' setting is been selected, meaning that only the impact of each competence area on the 'OR Int Discretized' is modeled.



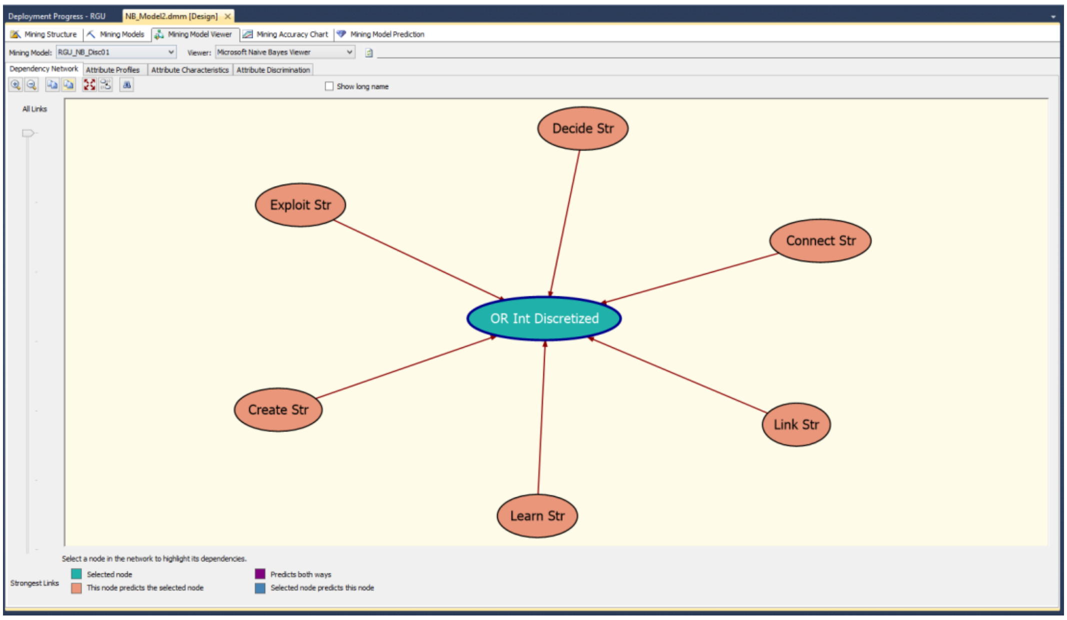
Structure	RGU_NB_Disc01
	Microsoft_Naive_Bayes
Connect Str	Input
Create Str	Input
Decide Str	Input
Exploit Str	Input
IP	Key
Learn Str	Input
Link Str	Input
OR Int Discretized	PredictOnly

Fig. 6.2.3.14: Mining model structure (called RGU_NB_Disc01), of the model NB_Model2

The following three figures (6.2.3.15, 6.2.3.16 and 6.2.3.17) present the dependency network that results from the construction of the mining model. Whereas Fig. 6.2.3.15 allows for viewing all six competence areas along with the role that each competence plays, the dependency probability parameter had to be set to 0.01 in order for the network to display all nodes. (Setting the value of the parameter to zero produced no output.) The resultant dependency network for NB_Model2 differs significantly from the one generated for NB_Model1 (RQ #3). In the illustration of the dependency network displayed in Fig.6.2.3.15, it is clear that the resultant model uses all competence areas to predict OR ('OR IntDiscretized') and 'OR Int Discretized' is the only attribute being predicted. The illustration in Fig. 6.2.3.16 shows exactly the same results as in the previous figure mining model but with the 'relationship' level set to 0.51 (MINIMUM_DEPENDENCY_PROBABILITY = 0.51. No other parameters of the model were changed from the default values provided by the system.). With the parameter set to a value of 0.51, it can be seen which nodes (attributes) in the diagram generate the largest impact on the model's outcome: determining 'OR Int Discretized'. Those attributes are Create Str, Learn Str and Link Str (RQ #3).

Finally, the illustration in Fig. 6.2.3.17 shows, represented by an edge, the strongest relationship in the model, which exists between 'Create Str' and 'OR Int Discretized,' indicating that the questionnaire replies in the Create Area tend to have the strongest correlation with the 'OR Int Discretized' (RQ #3). (The Create competence area was the first area in the questionnaire, consisting of eight questions relating to knowledge creation, acquisition and exploration. The questions asked attempted to identify any gaps in knowledge or knowledge-related processes that could provide insight into competitiveness.)

In the NB_Model2, using the Naïve Bayes algorithm when inspecting the resultant dependency network, it can be seen which specific input variables (competence areas: Create, Learn and Link) impact OR the most (RQ #3). The variation in the display in the form of the number of input variables is due to the setting of the algorithm's parameter asking for the minimum probability dependency to consider (with the value of this parameter set higher, the display includes only the competence areas impacting OR to the greatest extent). In addition to the display of the most influential competences, the dependency network showed, in the form of links, the strongest relations between the competence area and the OR. For the NB_Model2 containing very limited amount of data, the strongest link, or the most influential competence area (providing an answer to RQ #3), was the Create competence. Determination of the strongest links can provide a viewing context for DM-based analysis (RQ #3).

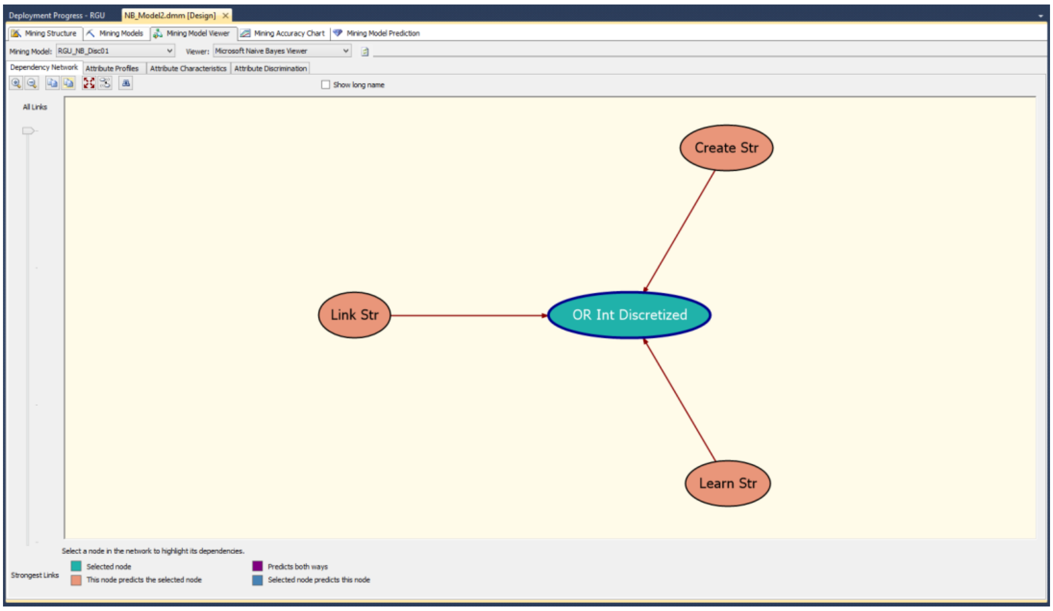


Mining Model: RGU_NB_Disc01

Viewer: Microsoft Naive Bayes Viewer

Select a node in the network to highlight its dependencies.

Fig. 6.2.3.15: Dependency network with the parameter MINIMUM_DEPENDENCY_PROBABILITY set to 0.01



Mining Model: RGU_NB_Disc01

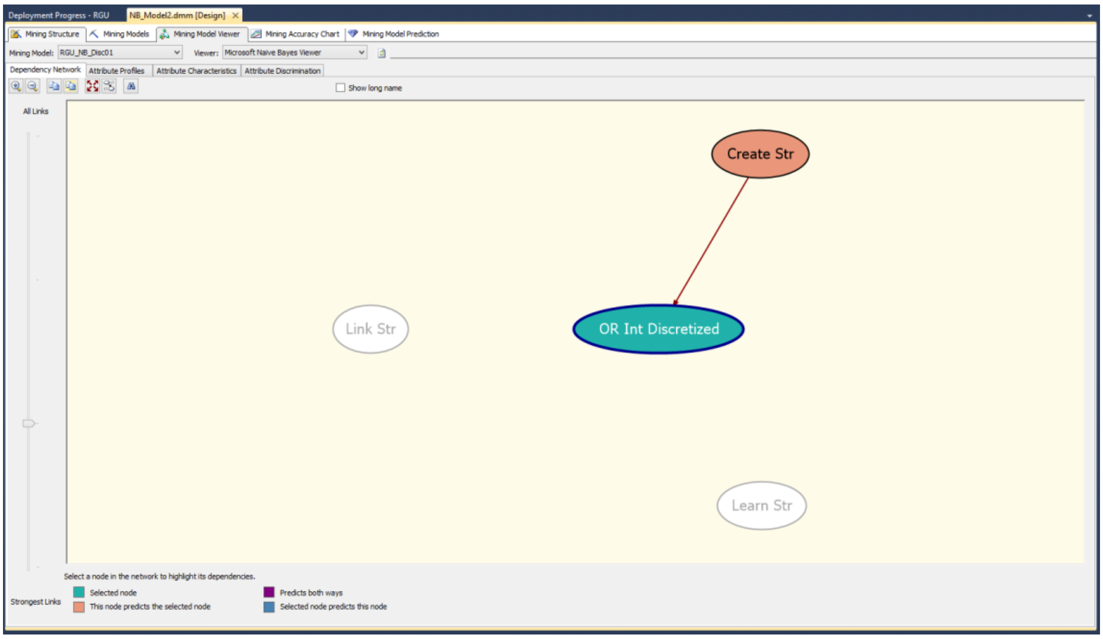
Viewer: Microsoft Naive Bayes Viewer

Select a node in the network to highlight its dependencies.

Fig. 6.2.3.16: Dependency network with the parameter MINIMUM_DEPENDENCY_PROBABILITY set to 0.51

The attribute profiles section presented in Fig. 6.2.3.18 shows how each one of the six competence areas corresponds to the output attribute (called 'Predictable' on the illustration and set to the only predictable attribute in the model, 'OR Int Digitized'). With the value of the parameter, for the remainder of this chapter, set to 0.51, only three attributes 'qualified' to be displayed.

(Clearly, the value of the MINIMUM_DEPENDENCY_PROBABILITY parameter can be set to a lower value, but, for the purpose of this research, the interest lies in illustrating how the most meaningful results can be obtained from the DM model. With that in mind, only attributes meeting the set threshold values, which are therefore more likely to be predictive of the output attribute, are discussed.)



Mining Model: RGU_NB_Disc01

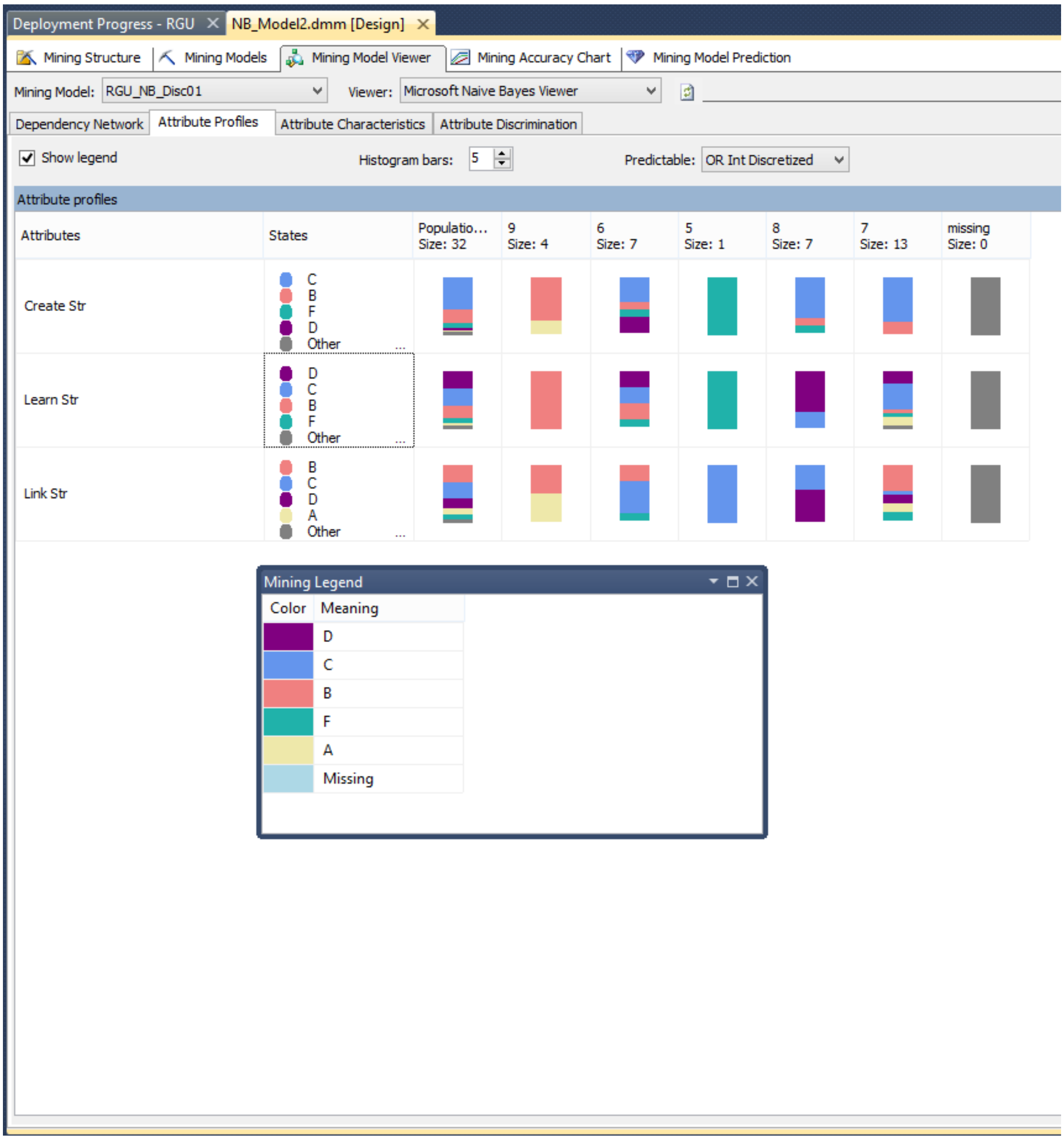
Viewer: Microsoft Naive Bayes Viewer

Select a node in the network to highlight its dependencies.

Fig. 6.2.3.17: Dependency network with the parameter MINIMUM_DEPENDENCY_PROBABILITY set to 0.51 and with the strongest link shown

As stated by MacLennan et al. (2009, pg. 234), '[s]etting the MINIMUM_DEPENDENCY_PROBABILITY parameter does not impact model training or prediction. Instead, it allows you to reduce the amount of content returned by the server from the content queries.'

With no missing values detected in the model and with the five distinct values ('A' to 'F'), the number of histograms for the inspection of attribute profiles has been set to five, to match each possible value within an input attribute. The resultant profiles are presented in Fig. 6.2.3.18. (Note that the states 'A' to 'F' are listed in decreasing order of occurrence in the population.)



Mining Model: RGU_NB_Disc01

Viewer: Microsoft Naive Bayes Viewer

☑ Show legend

Histogram bars: 5

Predictable: OR Int Discretized

Attributes	States	Populatio... Size: 32	9 Size: 4	6 Size: 7	5 Size: 1	8 Size: 7	7 Size: 13	missing Size: 0
Create Str	C B F D Other ...							
Learn Str	D C B F Other ...							
Link Str	B C D A Other ...							

Color	Meaning
🟪	D
🟦	C
🟥	B
🟩	F
🟨	A
🟦	Missing

Fig. 6.2.3.18: Profiles of the attributes of NB_Model2

Inspecting the attribute profiles (Fig. 6.2.3.18) indicates that thirty-two is the size of the population of the model and is, as can be seen, the largest group, made up out of thirteen elements that constitute a value of seven for the 'OR Int Discretized' output attribute.

A visual inspection of data in the attribute profiles, pictured in Fig. 6.2.3.18, reveals some key characteristics about each of the three attributes and various values of output, the predictable variable (RQ #3).

Inspecting the 'OR Int Discretized' across the input variables (vertically) indicates the following:

- Unsurprisingly, all three input variables show a high concentration of 'B' and 'A' states for the value of '9' of 'OR Int Discretized'. The 'Create Str' captures all 'A' states for 'OR Int Discretized' = 9. Somewhat surprising is the lack of presence of state 'A' in the composition of the input variable 'Learn Str.'

  As expected, there was a high concentration of state = 'F' in the lowest scoring (equal to five) value of the output variable; and
- A surprisingly diverse composition, in terms of state values, for 'OR Int Discretized' = '7', the largest group. The composition consists of many states with values of 'C' and 'B' but also 'A'.

Inspecting each of the input variables across the values of 'OR Int Discretized' (horizontally) indicates the following:

- 'Create Str', the strongest predictor in the model, appears to have 'expected' composition in the 'OR Int Discretized' = '9' category, containing only 'B' and 'A' states, with the A state not appearing in any other group. The composition of the '6' and '5' group is also somewhat expected as it contains state values of 'F', 'D' and 'C'. Somewhat surprising is the composition of the groups '8' and '7'. Group '7' contains 'higher valued' states than those of group '8' states, yet it is a category lower;
- 'Learn Str' appears to have the expected composition for groups '9' and '5,' as group '9' is entirely made up of the 'B' state and group '5' of 'F' state. Group '6' has an equal number of 'D', 'C' and 'B' elements, but it also has a significant number of 'F' elements. Again, the most interesting is the composition of the '8' and '7' groups. Perhaps counter intuitively, group '8,' which is composed of two states, has a majority of members from the 'D' state and a third of that of the 'C' state. Yet, the lower ranked group '7,' which is largely composed of the 'C' state values, has half as many as 'C' and 'D' values. It also has a trace of 'F' and 'B' values; and
- 'Link Str' has a clearly defined '9' group that is composed half of the 'A' and half of the 'B' state values, therefore scoring 'at the top'. The composition of the remaining groups appears to be somewhat less clear than in the case of the previous two input variables. The last group,

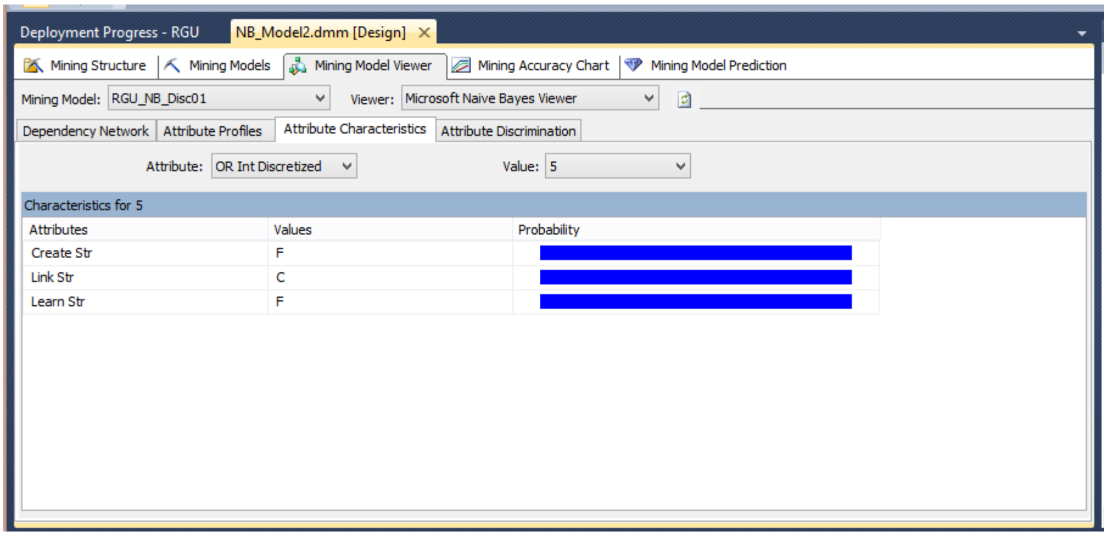
group ‘5,’ consists only of ‘C’ valued states (granted, a single element in that group), and group ‘6’ is composed mostly out of ‘C’ and ’B’ valued states, with very small ‘F’ valued elements. However, the much higher-ranked group ‘8’ consists only of two types of states: the larger ‘D’ and smaller ‘C’. Group ‘7’ appears to be a composition of all valued states, including equal amounts of values of states ‘A’, ‘D’ and ‘F’.

The attribute profiles section of NB_Model2 considered the most important competence areas (those that are most important to the NB algorithm based on the value of the minimum probability dependency variable) and their composition for every value of the predicted OR variable. This tool allows visual inspection of the outcome and answering questions such as ‘what is the composition of the competence area, in terms of the input variable, showing the greatest/smallest OR?’ when looking at the attribute profiles (like the one shown in Fig. 6.2.3.18) in row-wise fashion. It also allows looking at the results in column-wise fashion, which makes it possible to obtain answers to questions such as ‘what is the composition of the group holding a certain value of OR in terms of the most significant competence areas?’ Clearly, the entire population used in the analysis can be also inspected for either distribution of OR values within specified input variable (referred to in Fig. 6.2.3.18 as ‘States’) or the presence of all of the various ‘States’ in all variables. In short, the functionality of the attribute profiles allows for quick visualization of the composition of the competence area-OR pair of interest.

The attribute characteristics tab allows for building a deeper understanding, expressed as a probability, of the attributes present in a given group (‘9’ to ‘5’) of the predicted variable. As pointed by MacLennan et al. (2009, pg. 227), and already mentioned in this work, the two issues about the attribute characteristics tab that need to be kept in mind are 1) an attribute characteristic is not an implication of its predicting capabilities and, 2) the input values displayed are only those that satisfy the value of the parameter MINIMUM_DEPENDENCY_PROBABILITY (RQ #5).

Figures 6.2.3.19 and 6.2.3.20 illustrate the workings of the attribute characteristics tab with relation to the NB_Model2. In the case of predicting the output value of ‘5’ (Fig. 6.2.3.19), it can be seen (by placing the cursor on the blue probability line and reading the displayed value) that there is a 100% probability of detecting the state value of 'F' in 'Learn Str' and 'Create Str' and the state value of 'C' in 'Link Str' being associated with an output value of '5'.

Fig. 6.2.3.20 shows the probability of the presence of various state values for the key three input variables when the value of the predicted variable equals '9'. It can be seen that, in that case, the probability of state value 'B' for 'Learn Str' is the largest and is equal to 100%. The probability for state value 'B' of 'Create Str' being present is 75%. The probability of state value 'A' and 'B' for 'Link Str' is, in both cases, 50%. Finally, the probability of 'Create Str' taking on value 'A' is only 25%. In the same fashion, probabilities of other state values can be inspected for all other values of the output variable 'OR Int Str' by selecting an appropriate value from the drop-down box labeled 'Value'.



Mining Model: RGU_NB_Disc01 Viewer: Microsoft Naive Bayes Viewer

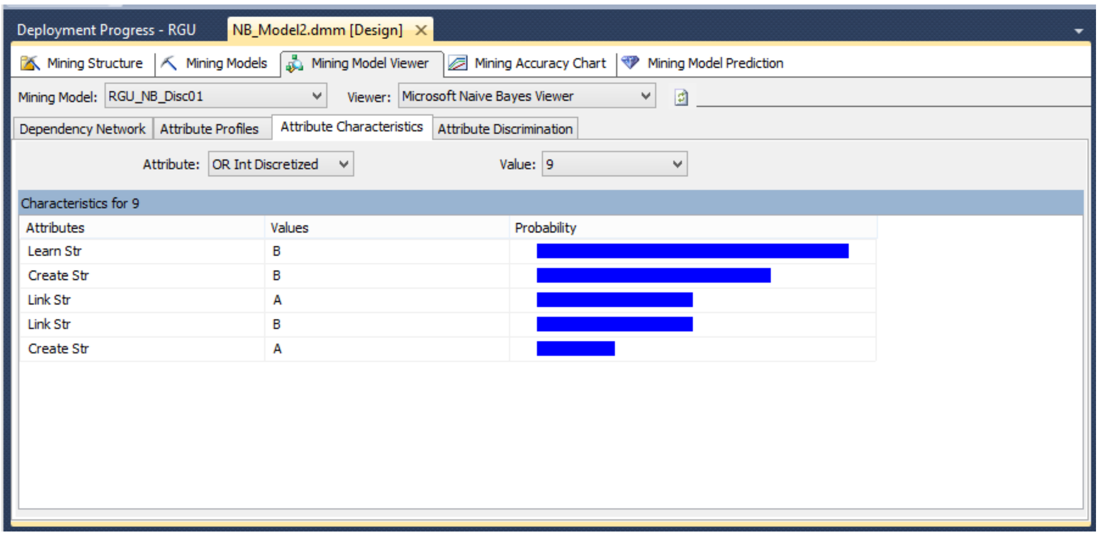
Attribute: OR Int Discretized Value: 5

Characteristics for 5

Attributes	Values	Probability
Create Str	F	█
Link Str	C	█
Learn Str	F	█

Fig. 6.2.3.19: Attribute characteristics for the value of '5' of the output variable 'OR Int Discretized'

While information displayed in the attribute characteristics is not implication of the model's predicting capabilities and the display is also affected, it shows the probability of the presence of certain OR value for the key variables (the Learn, Create and Link competence areas), the presence of which is influenced by setting the minimum dependency probability parameter appropriately (currently 0.51). As shown in figures 6.2.3.19 and 6.2.3.20 and earlier in figures 6.2.3.5 and 6.2.5.6, attribute characteristics provide information about the probability of the output variable obtaining certain values for the key competence areas. In the specific case investigated and presented in Fig. 6.2.3.20, the highest 'OR Score' of 9, on the scale of 5 to 9, had a 100% probability of being 'B' as a part of the Learn competence. Therefore, using the attribute characteristics functionality of SQL Server, it is possible to investigate the composition of the output variable in terms of probabilities of the values of the input variables (RQ #3).



Mining Model: RGU_NB_Disc01

Viewer: Microsoft Naive Bayes Viewer

Attribute: OR Int Discretized

Value: 9

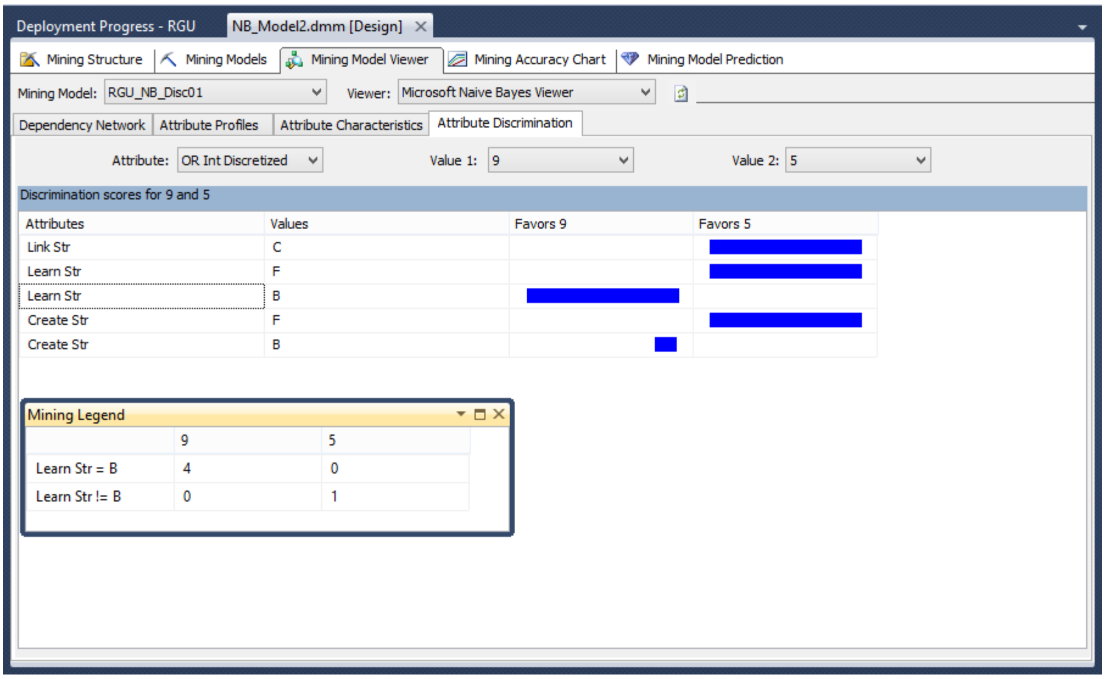
Characteristics for 9

Attributes	Values	Probability
Learn Str	B	
Create Str	B	
Link Str	A	
Link Str	B	
Create Str	A	

Fig. 6.2.3.20: Attribute characteristics for the value of '5' of the output variable 'OR Int Discretized'

Whereas the attribute characteristics tab allows for the inspection of a specific input attribute value, the attribute discrimination tab allows one to examine NB_Model2 and ask the question 'what is the difference between KM processes that favor output value "9" versus "5"?' The data presented on this tab, similarly to the data presented on the tabs already discussed, also directly supports the aims and objectives of this research. Specifically, support for the goals of this research comes in terms of understanding the relationships between KM and OR, how each competence area impacts OR and determining if, perhaps, some of the competence areas are more important than others. Fig. 6.2.3.21 presents the discrimination between the predicted values of '9' (the best) and '5' (the worst). Also, per the suggestion of MacLennan et al. (2009, pg. 228), the second type of discrimination, which seeks to determine unique discrimination scores for output '9' and all other output states, is presented (in Fig. 6.2.3.22). Clearly, many other viewing presentations are possible, but, given the limited input data and the focus of this research, they are omitted.

Fig 6.2.3.21 displays which factors favor the outcome value of ‘9’ and which the value of ‘5’. Disregarding the minimal data support presented in the mining legend, as the actual results are not the focus of this research, it can be seen that the value of ‘B’ of ‘Learn Str’ 100% favors the output value of ‘9’. The value of ‘B’ of the ‘Create Str’ favors, with a probability of 15 % (the percentage is visible upon placing the cursor on the blue measurement line), the output value ‘9’. Favoring the output value of ‘5’, with a probability of 100%, are the input state values of ‘F’ for both ‘Create Str’ and ‘Learn Str’ and the input value of ‘C’ for the ‘Link Str’ variable.



Mining Model: RGU_NB_Disc01 Viewer: Microsoft Naive Bayes Viewer

Attribute: OR Int Discretized Value 1: 9 Value 2: 5

Discrimination scores for 9 and 5

Attributes	Values	Favors 9	Favors 5
Link Str	C		▬
Learn Str	F		▬
Learn Str	B	▬	
Create Str	F		▬
Create Str	B	▬	

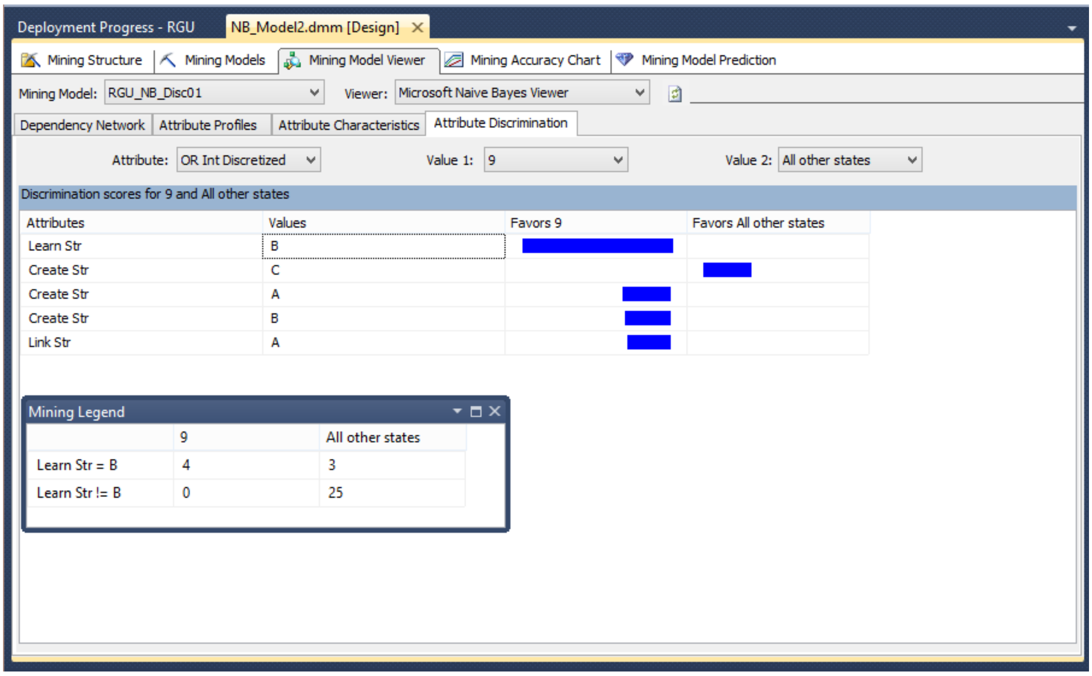
Mining Legend

	9	5
Learn Str = B	4	0
Learn Str != B	0	1

Fig. 6.2.3.21: Model NB_Model2, discrimination scores for ‘9’ and ‘5, with the ’Learn Str’ input variable selected

Fig. 6.2.3.22 shows the discrimination scores for ‘9’ and all other states for NB_Model2. In the figure, it can be seen that ‘Learn Str’ value of ‘B’ favors (100%) being found in the output value of ‘9’ only. Other values, ‘A’ for ‘Create Str’ and ‘Link Str’ and ‘B’ for ‘Create Str’, favor output value ‘9’ but are not exclusively found in ‘9,’ as was the case with the input value of ‘B’ of the ‘Learn Str’. Interesting, although not displayed, is the fact that, when evaluating the output value of ‘5’ and all other states, the input values of ‘F’ for ‘Create Str’ and ‘Learn Str’ exclusively favor the output ‘5’.

The attribute discrimination tab of NB_Model2 illustrates some of the most interesting findings of this work. With the help of the attribute discrimination functionality, it is possible to compare the most resilient organizations (those with ‘OR Score’ = 9) with the least resilient ones (those with ‘OR Score’ = 5) (RQ # 3). Similarly to the other SQL Server aspects already described, the attribute discrimination tool provides, as a percentage, information what value of the input attribute favors the most or least resilient score.



Mining Model: RGU_NB_Disc01

Viewer: Microsoft Naive Bayes Viewer

Attribute: OR Int Discretized

Value 1: 9

Value 2: All other states

Discrimination scores for 9 and All other states

Attributes	Values	Favors 9	Favors All other states
Learn Str	B	█	
Create Str	C		█
Create Str	A	█	
Create Str	B	█	
Link Str	A	█	

	9	All other states
Learn Str = B	4	3
Learn Str != B	0	25

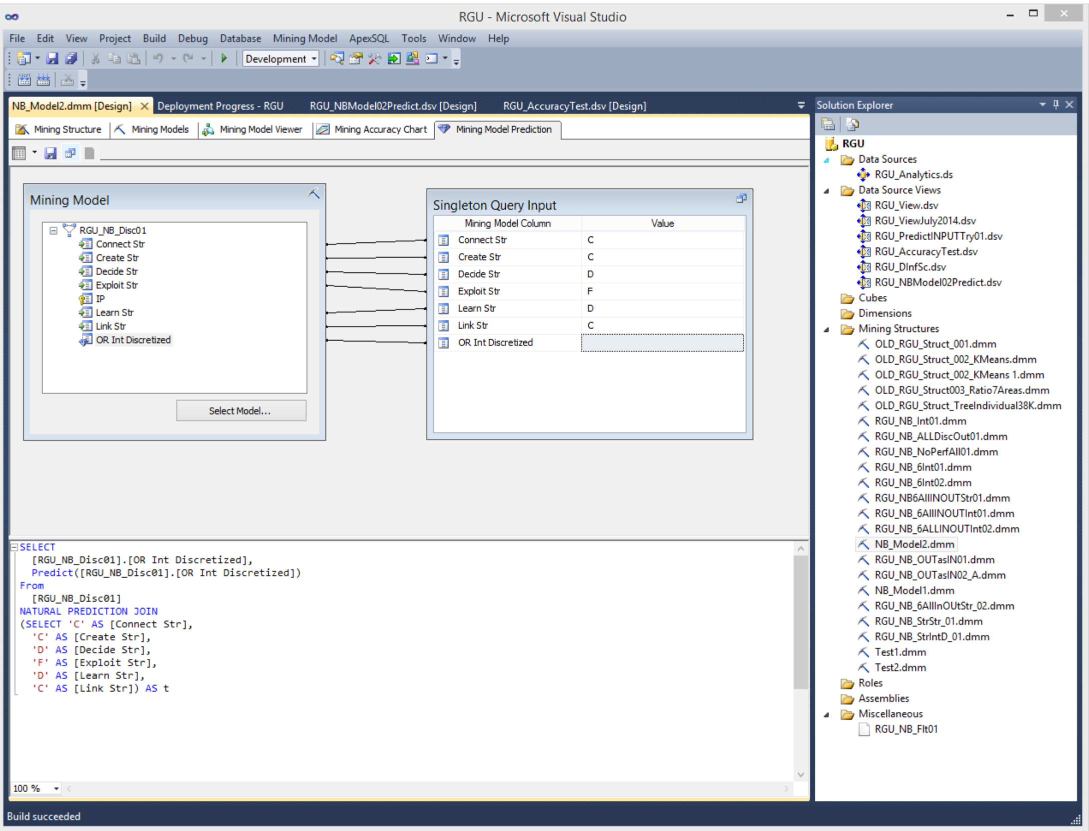
Fig. 6.2.3.22: Model NB_Model2, discrimination scores for ‘9’ and ‘all other scores’, with the ’Learn Str’ input variable selected

Comparison of specific values of the ‘OR Score’, presented in Fig. 6.2.3.22, with all other values of the output variable makes it possible to capture uniquely favored values for a specific ‘OR Score’, if such unique values exist.

With the predictive ability being the next topic, the focus for the remainder of this section is on demonstrating the possible practical application of the model: predicting the ‘OR Score’ for an organization.

For the purpose of making a prediction, a questionnaire reply was randomly selected from all of the replies. The company selected was a medical supply firm, with IP = 13. The values for all input variables were entered into the

‘Query Input’ window (see Fig. 6.2.3.23), and the resulting query is shown in the lower portion of the window.



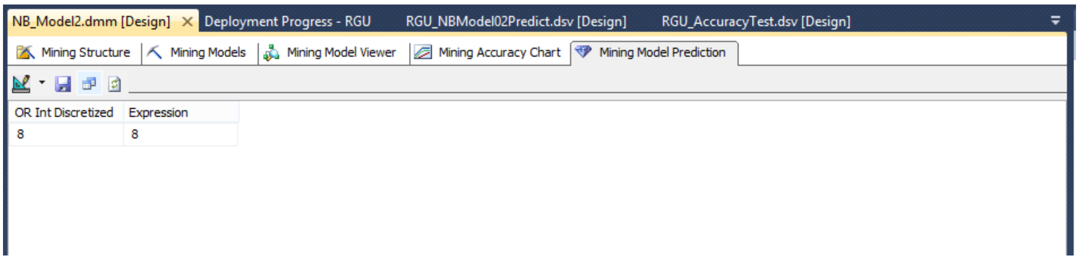
Singleton Query Input

Mining Model Column	Value
Connect Str	C
Create Str	C
Decide Str	D
Exploit Str	F
Learn Str	D
Link Str	C
OR Int Discretized	

```
SELECT
  [RGU_NB_Disc01].[OR Int Discretized],
  Predict([RGU_NB_Disc01].[OR Int Discretized])
From
  [RGU_NB_Disc01]
NATURAL PREDICTION JOIN
(SELECT 'C' AS [Connect Str],
  'C' AS [Create Str],
  'D' AS [Decide Str],
  'F' AS [Exploit Str],
  'D' AS [Learn Str],
  'C' AS [Link Str]) AS t
```

Fig. 6.2.3.23: NB_Model2 prediction

Executing the constructed predictive mining model results in the following outcome (Fig. 6.2.3.24): As illustrated in the figure, the predicted OR Score for the company IP = 13 is ‘OR Int Discretized’ = ‘8’.



OR Int Discretized	Expression
8	8

Fig. 6.2.3.24: Output of the prediction based on NB_Model2 (the OR score for IP = 13)

One of the most practical outcomes of this research is the ability to predict the ‘OR Score’ of an organization based on questionnaire replies that have been further processed in order to populate the ‘OR Scores’ for each competence area.

Upon calculating the 'OR Score', the other tools described above can be engaged to determine the reason for the score or areas for improvement. Moreover, the prediction tool makes it possible to run simulations to investigate various possible scenarios.

Finally, when the prediction ability of the model was tested with a data set that was randomly selected for this purpose, the resulting 'OR Score' prediction was 8 (with '9' being the highest and '5' the lowest scores in the model's data set). This method of prediction could be applied to any company that provides replies to the questions in the questionnaire used in this research.

In summary, the easy-to-construct NB_Model2 model provided very insightful information regarding which competence areas lead to an organization being resilient (RQ #3). The clear graphical presentation facilitates easy comprehension of the findings. However, a widely known issue of the NB-based algorithms is that it considers a single competence area at a time; thus, its impact on OR could not be confirmed (due to limited amount of data) as a deficiency (RQ #5).

6.3 Data Mining: Clustering

While the prior section (6.2) examined the most influential KM processes (which are composed of numerous activities and combined into a logical group called the competence area, as defined in Section 4.6.2), this section considers a more granular level: a single activity within a competence area. Such a single activity corresponds to a single question in a questionnaire. This section discusses the determination of key activities within the KM processes responsible for OR; it also discusses an alternative to the competence area approach to grouping KM-activities (with a KM activity, for the purpose of this discussion, being an activity asked in a single question in the questionnaire).

As defined by Han et al. (2012, pg. 414) '[c]luster analysis or simply clustering is the process of partitioning a set of data objects (or observations) into subsets. Each subset is a cluster, such that the objects in a cluster are similar to one another, yet dissimilar to objects in other clusters.' The primary purpose of using the clustering algorithm is to discover previously unknown groupings within data. In this research, the clustering algorithm will be applied to individual questions contained in the questionnaire in order to obtain an alternative to the six competences grouping of input parameters.

The clustering algorithm is an example of an unsupervised learning type of algorithm, as opposed to the NB algorithms presented in the previous section, where the class label is not provided. That is, when the algorithm is employed, the desired characteristics of the resultant group (or cluster/segment) are not specified. (The term segmentation is, in the industry, synonymous with the term classification.) Rather, the clustering algorithm derives the clusters/segments on its own, according to rules programmed into it, which are briefly explained below.

As stated by Larose and Larose (2015, pg. 523), 'the clustering task does not try to classify, estimate, or predict the value of a target variable. Instead, clustering algorithms seek to segment the entire data set into relatively homogenous subgroups or clusters, where the similarity of the records within the cluster is maximized, and the similarity to records outside this cluster is minimized.'

One of the key aspects of the clustering algorithm is the method used to assign of an element to a cluster. The clustering algorithm offered by Microsoft Corporation uses two distinct methods for assigning an element to a cluster: the K-means method and the expectation maximization (EM) method. As described by Witten et al. (2011, pg. 139), K-means assigns cluster membership by distance, where an element belongs to the cluster with closest center element (using a simple Euclidean distance as a measure). Once all elements have been assigned to the clusters, the center of a given cluster is moved to the mean of all elements that make up the given cluster (hence the name: K-means). With the K-means method, an element can belong to, at most, one cluster.

The EM method, according to MacLennan et al. (2009, pg. 311), uses probabilistic measure (instead of strict distance formula), so, instead of computing a distance to the center of a cluster, it uses a bell curve (with a mean and standard deviation) for each dimension. Then, an element falling on a bell curve is assigned to a cluster with a certain probability. With the EM method, an element can belong to more than one cluster (clusters can have common elements) because the bell curves can (and often do) overlap between the clusters.

It is also worthwhile to mention that the clustering technique is an iterative method that requires numerous iterations with the training data set in order to arrive at the segmentation.

As stated by Larose and Larose (2015, pg. 16) 'clustering is often performed as a preliminary step in a data mining process, with the resulting clusters being used as further inputs into a different technique downstream, such as neural networks.' With regard to this research, the results of the clustering could have been used as a grouping of questionnaire answers in place of the six competence areas – this is discussed further in Section 6.3.3.

6.3.1 Clustering Preliminaries

Similarly to other DM algorithms, the clustering algorithm has a set of conditions that must be satisfied in order for the algorithm to produce meaningful results.

Han et al. (2012, pg. 484) identify some of the prerequisites in relation to the use of the clustering algorithm:

- Referred to as the 'clustering tendency assessment,' the goal of the assessment is to determine whether a given data set has a non-random structure, which may lead to the creation of meaningful clusters. That is, clustering requires a non-uniform distribution of data (RQ #5); and
- The number of (output) clusters required to ensure proper granularity of cluster analysis must be determined (RQ #5). This, according to Han et al., and other writers (Abbott, 2012, pg. 185; de Ville, 2001, pg. 154) and practitioners, is no easy task, as it tends to require a tradeoff between compressibility (aggregated values) and accuracy (the smallest distances possible between a data element and a cluster's center). For the purpose of this research, a model is built using two approaches: allowing the system to determine the optimal number of clusters and using six clusters to correspond to the six competence areas. With the focus of this research being on methodology instead of actual numerical results, simple selection criteria for the selection of the number of clusters suffices.

Abbott (2014, pg. 183) provides additional requirements with regard to the data used by the algorithm:

- Reduce skew, if needed. (The largest absolute value of skewness encountered in the questionnaire's responses was 1.6 [in questions: Exploit_References, Decide_Condition, Link_Relationship and Link_Actively], with the majority of answers having an absolute value skewness score of less than 1.0. Should there be a need to make the data 'more normally distributed,' Larose and Larose [2015, pg. 35] suggest the use of data transformation tools such as the natural log transformation, the square root transformation or the inverse square root transformation);
- Include categorical variables only if necessary and after exploding them into dummy variables, as categorical variables are problematic in computing. (The data set used in this research for building the clustering model uses only continuous, not categorical, variables. Methods for addressing this issue were discussed in Section 5.4.2.2.)

  If needed, scale all inputs so that they are on the same scale. All input variables used in the clustering model are of the same scale (1 to 5 with 0 = N/A). Methods for addressing this issue were also discussed in Section 5.4.2.2.)

6.3.2 Clustering Model

The EM Method

Cluster_Model1 – the purpose of this model is to investigate the groupings and the characteristics of the grouping as replies to the questionnaire used in this research. All fifty-two questions (excluded were the sixteen questions related to performance and sixteen related to OR) were used as input variables, and while typically no prediction takes place in a clustering algorithm, the 'OR Integer' attribute was selected as 'Predict Only', which according to Microsoft's on-line documentation (Microsoft Corporation 'Microsoft Clustering Algorithm'), can be used to provide groupings based on the 'Predict Only' column. (As stated by MacLennan [2009, pg. 314], the clustering algorithms do not typically involve prediction. This special 'predictive' ability is Microsoft's interpretation of the algorithm.) The 'OR Integer' attribute used was of the 'Predict Only' variable type in order to investigate the resulting groupings in the context of that variable – this concept is discussed further when discussing the model results.

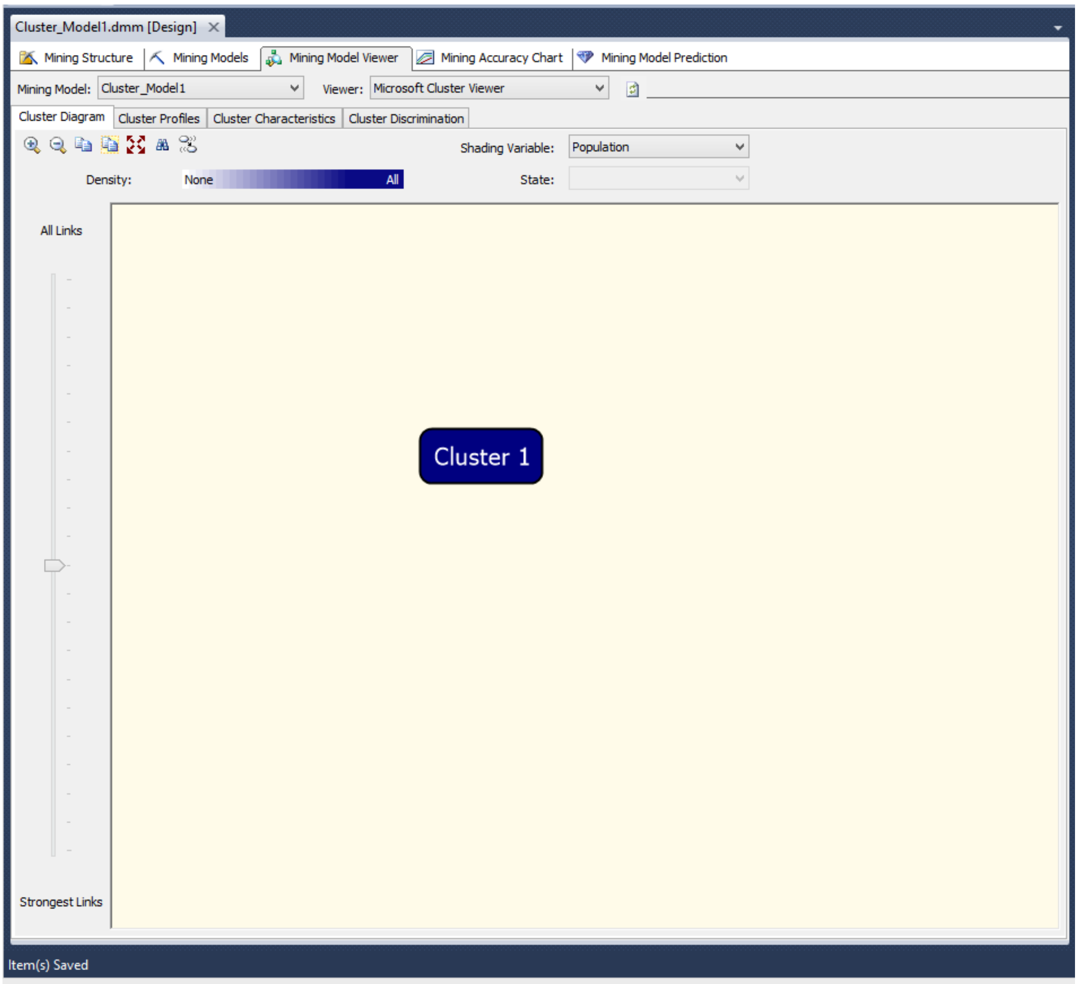
The structure used by the Cluster_Model1 model is presented pictorially in Fig. A8.19 and Fig. A8.20 in Appendix VIII and in Appendix XI.

The clustering model, Cluster_Model1, was constructed using the data source and 'RGU_DInfSc' data source view previously described when discussing the NB-based model. As before, the model was constructed using the DM wizard, using the default 30% of the data set for testing. Various aspects of the construction of the Cluster_Model1 are presented in Appendix VIII Fig. A8.18 – A8.22.

While a single model was constructed, the performance of the model is largely controlled by the parameters presented to it. The pictorial illustration of the algorithm's parameters is presented in Appendix VIII, Fig. A8.24 and in Appendix XII. Appendix XII presents the parameters that, as described by Microsoft's product literature (https://msdn.microsoft.com/en-us/library/cc280445(v=sql.110).aspx), control performance and accuracy of the resulting mining model along with a description of its use in this research.

In addition to the parameters described in Appendix XII, the modeling algorithm uses modeling flags (MODEL_EXISTENCE_ONLY and NOT NULL) that instruct the algorithm how to treat the null values when encountered. Since the input data used in this research contains no null values, there is no discussion of the impact of these modeling flags (RQ #5).

Initially, the Cluster_Model1 model was constructed with CLUSTER_COUNT parameter set = 0 (allowing the system to arrive at the optimal number of clusters), CLUSTERING_METHOD = 2 (non-scalable EM) and STOPPING_TOLERANCE = 4 (as the default value of 10 appeared too large for the 32-element input data set). The resulting model, most likely as a result of the very small input data set, produced a single cluster, as shown in Fig. 6.3.25. The resulting single-clustered model makes the model unsuitable for analysis, as there is simply no other cluster to compare the single existing cluster with, and all of the values of every attribute were put into a 'single bag' (RQ #5).



Mining Model: Cluster_Model1

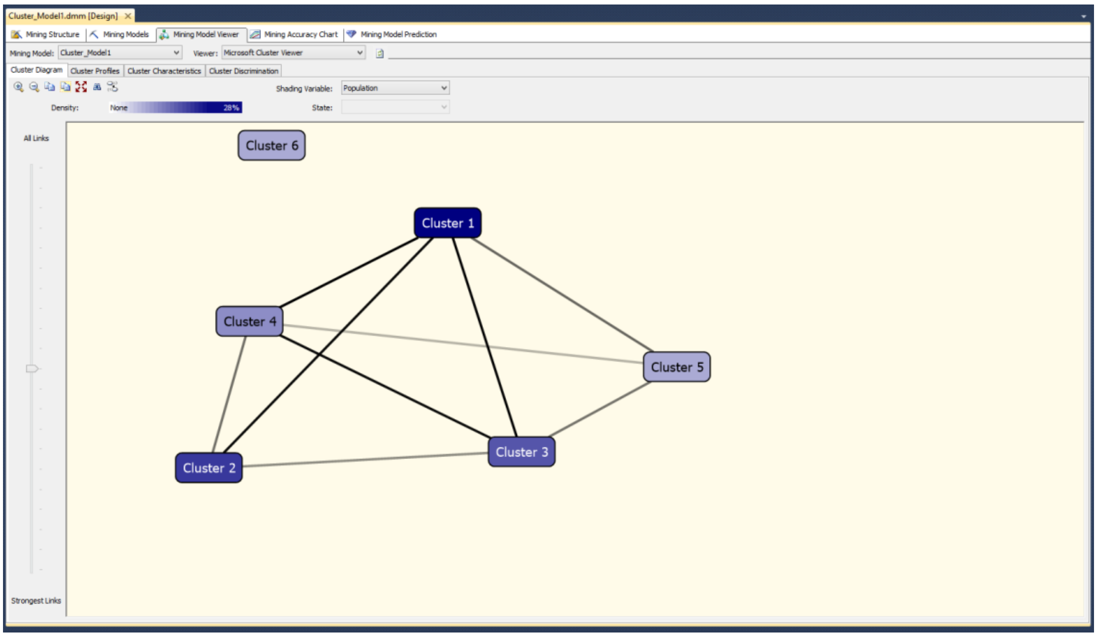
Viewer: Microsoft Cluster Viewer

Shading Variable: Population

Fig 6.3.25: Outcome of the modelling using clustering algorithm with the CLUSTER_COUNT parameter set to zero

When the value of the parameter CLUSTERING_METHOD was changed to 4, meaning that a non-scalable K-Means algorithm was used, while keeping the values of all other parameters constant, and a new model was created, the resulting model also contained a single cluster (RQ #5).

The next construction of the clustering model was controlled by the following settings of the parameters: CLUSTER_COUNT = 6 (to mimic the six competence areas as clusters), CLUSTERING_METHOD = 2 (non-scalable EM) and unchanged STOPPING_TOLERANCE = 4. The resulting model, showing the cluster diagram for the entire population, is presented in Fig. 6.3.26, below:


Mining Model: Cluster_Model1

Viewer: Microsoft Cluster Viewer

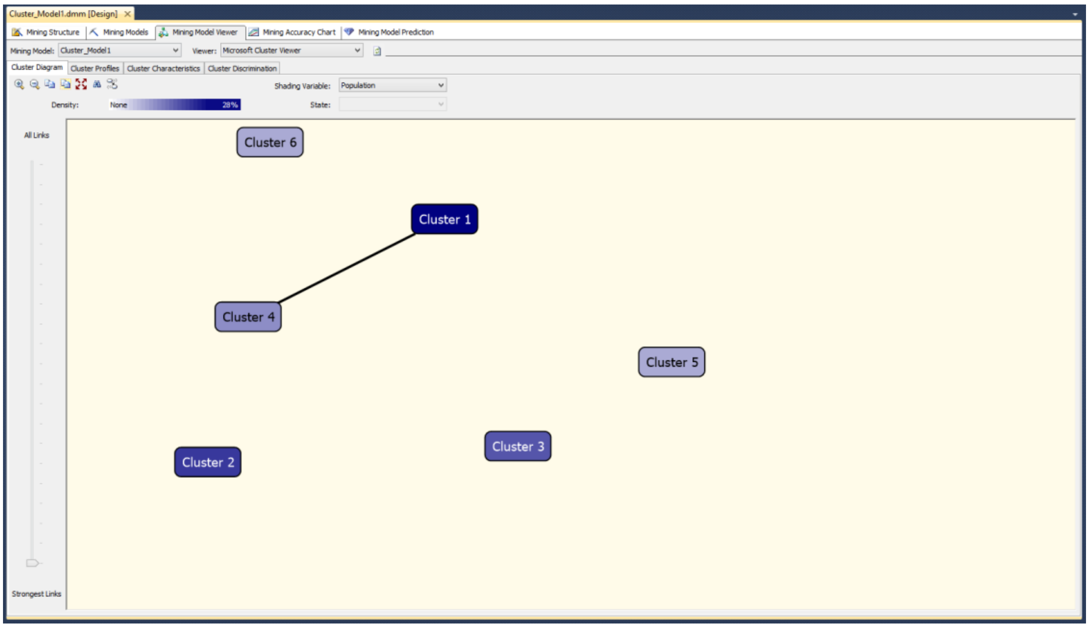
Shading Variable: Population

Density: None 28%

State:

Fig. 6.3.26: Clustering diagram for the entire population

The primary focus of the diagram presented in Fig. 6.3.26 is on the distribution of the data elements amongst the clusters. As indicated by the density option, Cluster 1 appears to contain about 28% of all input data. While looking at the 'Cluster Profiles', 'Cluster Characteristics', and 'Cluster Discrimination' tabs for the entire population is certainly worthwhile, the inspection of those tabs in the OR context is certainly more desirable. (It is possible to inspect the outcome of the clustering algorithm in terms of the 'output' variable because of Microsoft's implementation or clustering algorithm. By making an attribute 'Predict Only' in the mining model, it is possible to inspect the output of the clustering algorithm in terms of the output variable [RQ #3]. For this reason, a discussion of the output of the clustering algorithm takes place in the context of the 'OR Integer' output variable.)



Mining Model: Cluster_Model1

Viewer: Microsoft Cluster Viewer

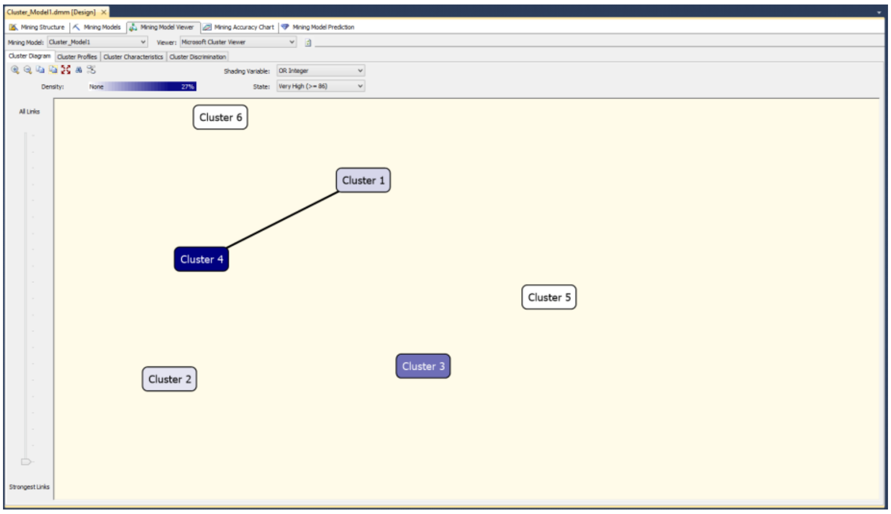
Shading Variable: Population

Density: None 28%

State:

Fig. 6.3.27: Clustering diagram for the entire population, with the strongest relationship shown

Inspecting the diagrams presented in Figures 6.3.28 and 6.3.29 (considering the fact that, in EM clustering, an element can be in more than one cluster – however, the sum of elements in each cluster still adds up to the size of the population) allows for identifying the strengths of the relationships between clusters and the high-level composition of the cluster (RQ #3). In the case of the diagram in Fig. 6.3.28, it indicates that Cluster 4 has a density of 27% of very high (greater than or equal to 86) values in the 'OR integer' (output) attribute.



Mining Model: Cluster_Model1

Viewer: Microsoft Cluster Viewer

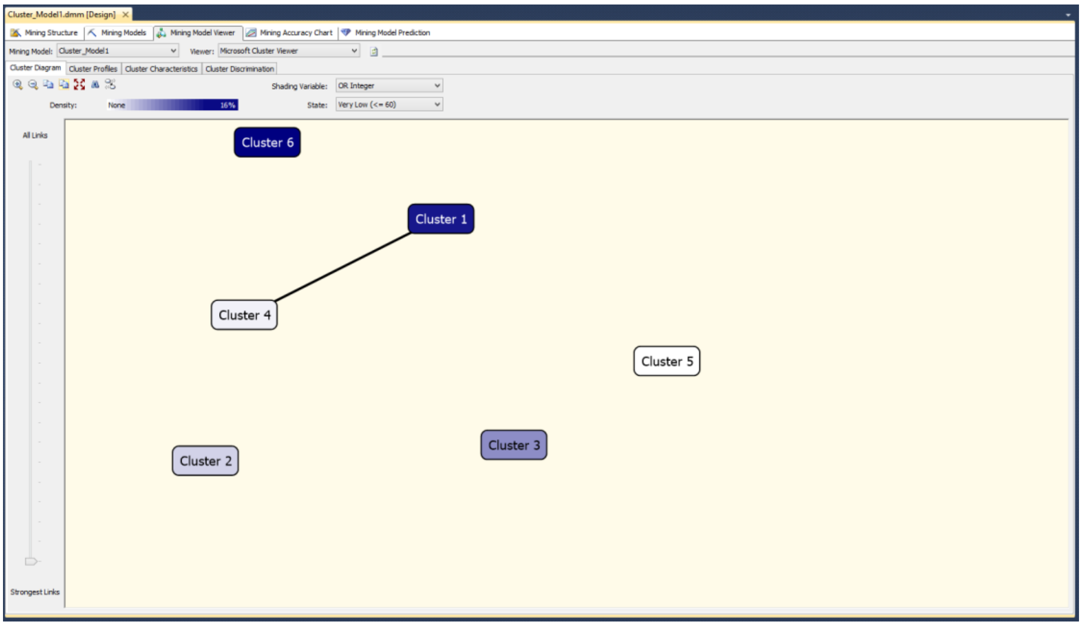
Shading Variable: OR Integer

Density: None 27%

State: Very High (>= 86)

Fig. 6.3.28: Clustering diagram showing the strongest relationship for the shading variable 'OR Integer' and value greater than or equal 86

An inspection of the diagram in Fig. 6.3.29 allows for the identification of a cluster, Cluster 6, with a density of 16% of values of 'OR Integer' less than or equal to 60 (very low values). A slightly lower concentration of very low values in the attribute 'OR Integer' is in Cluster 1, which has a slightly lighter color than Cluster 6.



Mining Model: Cluster_Model1

Viewer: Microsoft Cluster Viewer

Shading Variable: OR Integer

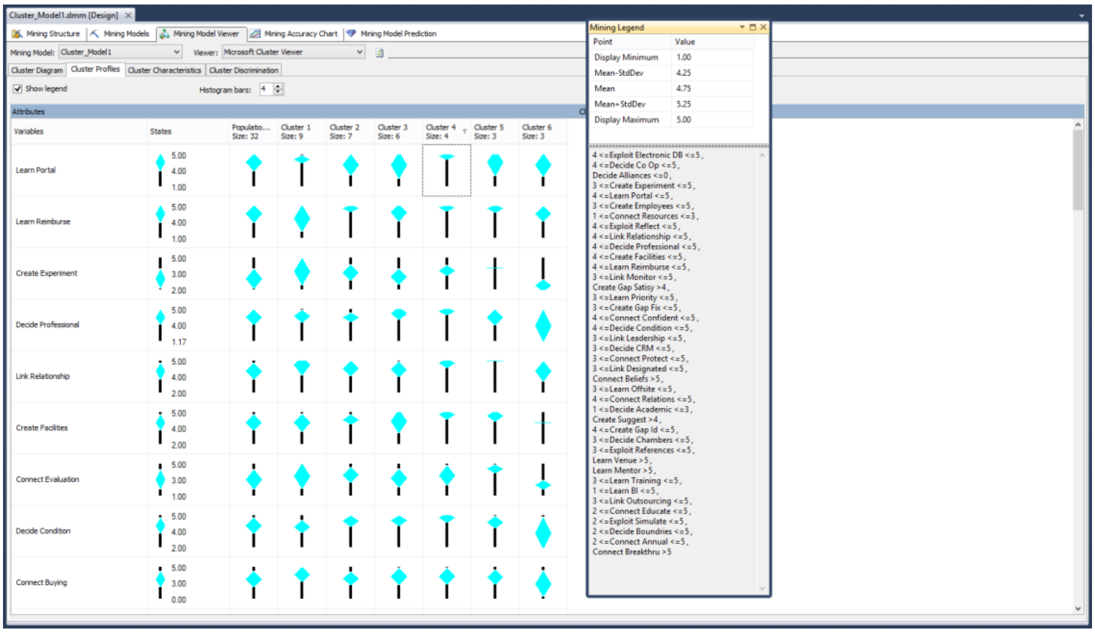
State: Very Low (<= 60)

Density: None 16%

Fig. 6.3.29: Clustering diagram showing the strongest relationship for the shading variable 'OR Integer' and value less than or equal to 60

The display of cluster profiles is similar to the attribute profiles of the NB models already presented, with small differences. The display of the continuous type of variables is no longer represented as a histogram, as was the case when inspecting the output of NB algorithms; rather, the output is represented by a diamond-shaped area that includes values that are associated with a given variable. The black line in the diamond chart represents the range values. The center of the diamond chart represents the mean for the variable; the width of the diamond represents the variance of the variable, implying that the thinner the diamond shape, the better the prediction.

Clicking on any one of the displayed cells (such as, in the case of Fig. 6.3.30, the intersection of the 'Learn Portal' question and Cluster 4) and viewing the mining legend reveals the variable's basic statistics (mean, standard deviation and alike). The lower portion of the mining legend displays information about the composition of the cluster that the selected element falls under (in the case of the element selected in Fig. 6.3.30, it is the information about Cluster 4). The columns, as was the case with the cluster description in the NB-based mode, in addition to the names of the segments, contain the count of elements within a given cluster.



Mining Model: Cluster_Model1

Viewer: Microsoft Cluster Viewer

☑ Show legend

Histogram bars: 4

Mining Legend

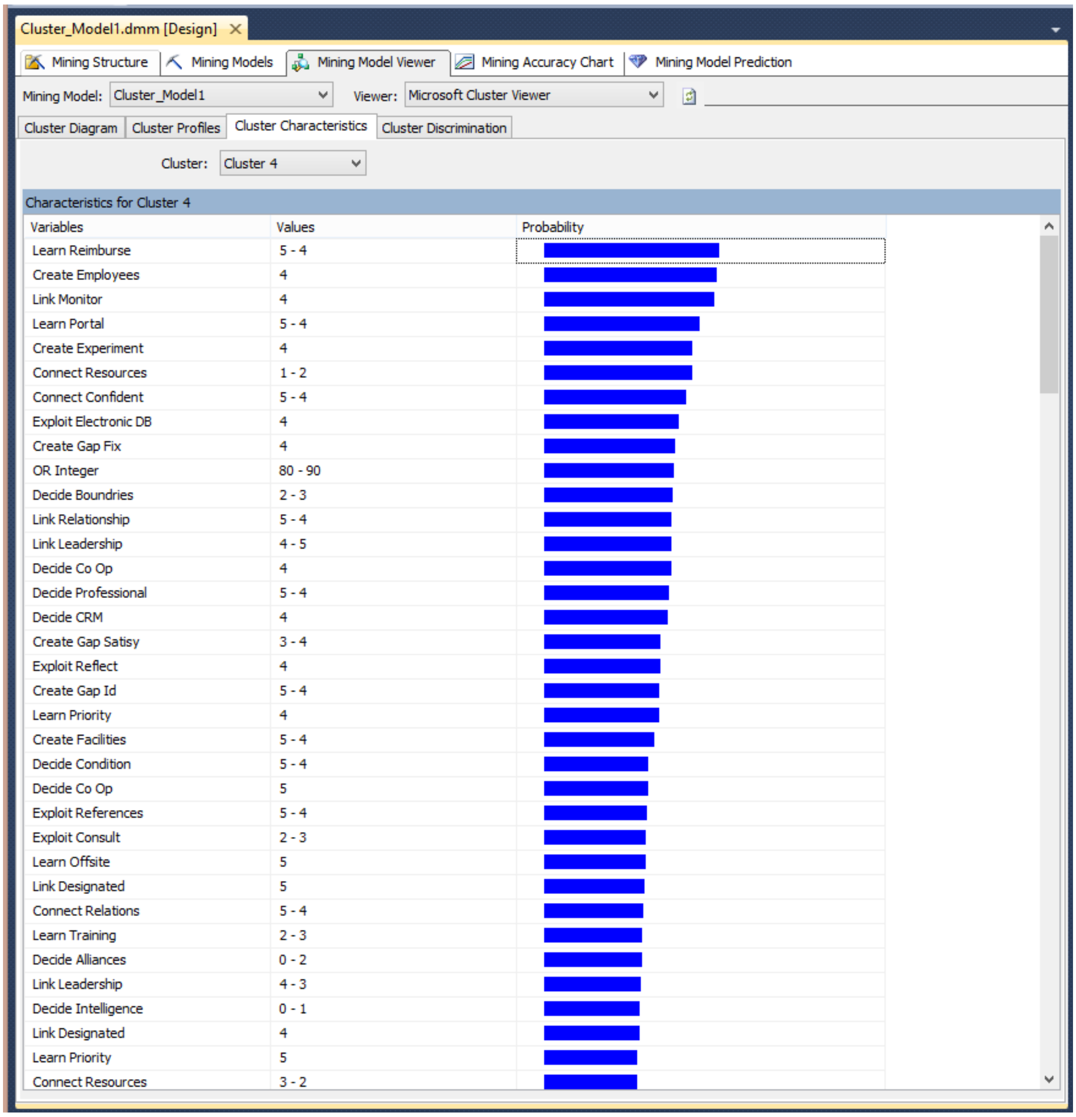
Point	Value
Display Minimum	1.00
Mean-StdDev	4.25
Mean	4.75
Mean+StdDev	5.25
Display Maximum	5.00

4 <=Exploit Electronic DB <=5,
4 <=Decide Co Op <=5,
Decide Alliances <=0,
3 <=Create Experiment <=5,
4 <=Learn Portal <=5,
3 <=Create Employees <=5,
1 <=Connect Resources <=3,
4 <=Exploit Reflect <=5,
4 <=Link Relationship <=5,
4 <=Decide Professional <=5,
4 <=Create Facilities <=5,
4 <=Learn Reimburse <=5,
3 <=Link Monitor <=5,
Create Gap Satisy >4,
3 <=Learn Priority <=5,
3 <=Create Gap Fix <=5,
4 <=Connect Confident <=5,
4 <=Decide Condition <=5,
3 <=Link Leadership <=5,
3 <=Decide CRM <=5,
3 <=Connect Protect <=5,
3 <=Link Designated <=5,
Connect Beliefs >5,
3 <=Learn Offsite <=5,
4 <=Connect Relations <=5,
1 <=Decide Academic <=3,
Create Suggest >4,
4 <=Create Gap Id <=5,
3 <=Decide Chambers <=5,
3 <=Exploit References <=5,
Learn Venue >5,
Learn Mentor >5,
3 <=Learn Training <=5,
1 <=Learn BI <=5,
3 <=Link Outsourcing <=5,
2 <=Connect Educate <=5,
2 <=Exploit Simulate <=5,
2 <=Decide Boundries <=5,
2 <=Connect Annual <=5,
Connect Breakthru >5

Fig. 6.3.30: Cluster profiles with mining legend shown

While the cluster profile tab provides a detailed graphical representation of the composition of each cluster in relation to the input variable, in the event of a large number of input variables (as was the case in this research), other tabs present on the mining model viewer display provide more focused and relevant information.

The cluster characteristics profiles pictured in figures 6.3.31 and 6.3.32 provide views of 'key for the cluster' variables on a more manageable level than that presented in Fig. 6.3.30. Considering the top three variables listed in Fig. 6.3.31 variables (only those variables with probabilities exceeding 50%, the random chance), it can be seen that, in Cluster 4 (the one with the highest 'OR Integer' score), the following variables have the greatest probability of being found in that cluster: 'Learn Reimburse' with values of 4-5 has a probability of 56%, 'Create Employees' with a value of 4 and a probability of 55% and 'Link Monitor' with a value of 4 and a probability of 54%.



Mining Model: Cluster_Model1

Viewer: Microsoft Cluster Viewer

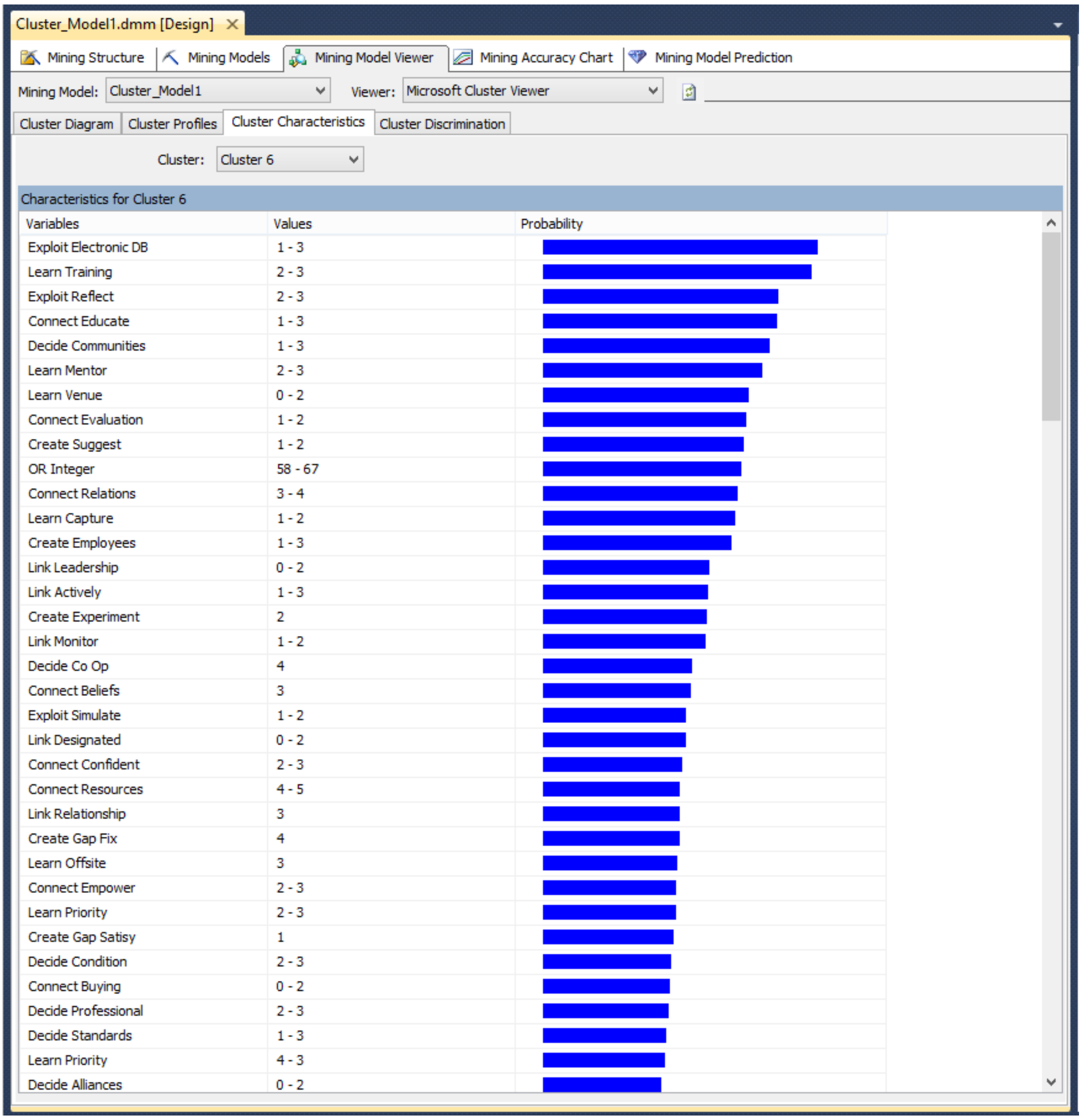
Cluster: Cluster 4

Characteristics for Cluster 4

Variables	Values	Probability
Learn Reimburse	5 - 4	
Create Employees	4	
Link Monitor	4	
Learn Portal	5 - 4	
Create Experiment	4	
Connect Resources	1 - 2	
Connect Confident	5 - 4	
Exploit Electronic DB	4	
Create Gap Fix	4	
OR Integer	80 - 90	
Decide Boundries	2 - 3	
Link Relationship	5 - 4	
Link Leadership	4 - 5	
Decide Co Op	4	
Decide Professional	5 - 4	
Decide CRM	4	
Create Gap Satisy	3 - 4	
Exploit Reflect	4	
Create Gap Id	5 - 4	
Learn Priority	4	
Create Facilities	5 - 4	
Decide Condition	5 - 4	
Decide Co Op	5	
Exploit References	5 - 4	
Exploit Consult	2 - 3	
Learn Offsite	5	
Link Designated	5	
Connect Relations	5 - 4	
Learn Training	2 - 3	
Decide Alliances	0 - 2	
Link Leadership	4 - 3	
Decide Intelligence	0 - 1	
Link Designated	4	
Learn Priority	5	
Connect Resources	3 - 2	

Fig. 6.3.31: Characteristics of Cluster 4 of the Cluster_Model1

Inspection of the composition of Cluster 6 (which has the lowest 'OR Integer' score), presented in Fig. 6.3.32, indicates more profound differences in probabilities than was the case with Cluster 4. It can be seen that organizations that do not exploit electronic databases (Exploit Electronic DB) with values between 1 and 3 have an 87% chance of being placed in Cluster 6. Also the firms that do not offer regular employee training (Learn Training) with values between 2 and 3 have an 85% chance of being a part of the Cluster 6. In all, there are sixteen variables (excluding the 'OR Integer' also listed as a variable) that have a greater than 50% chance of being placed in Cluster 6 (RQ #3).



Mining Model: Cluster_Model1

Viewer: Microsoft Cluster Viewer

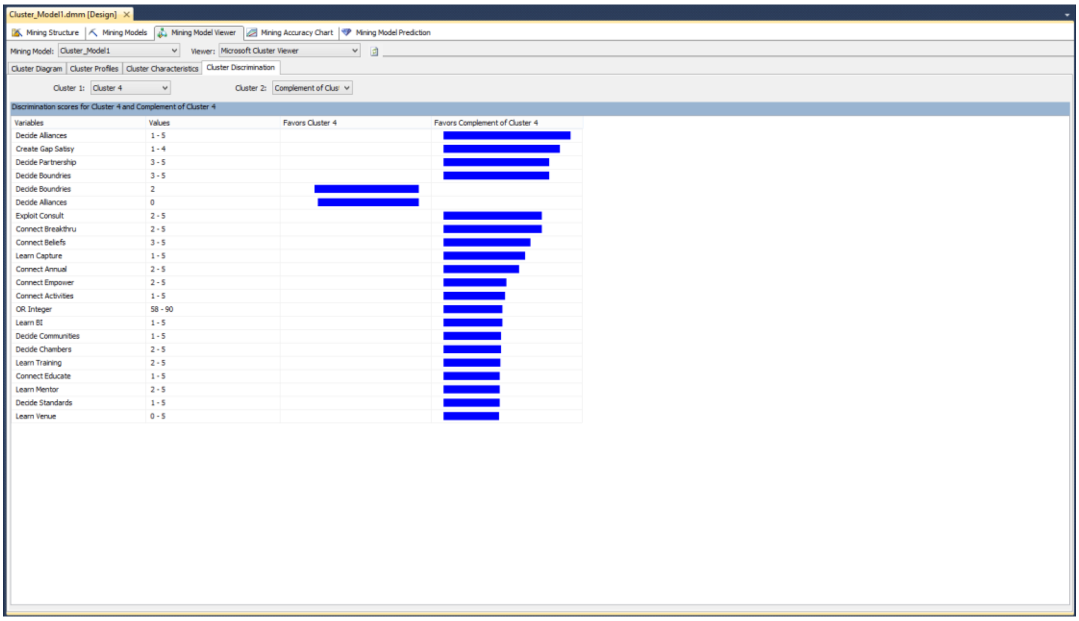
Cluster: Cluster 6

Characteristics for Cluster 6

Variables	Values	Probability
Exploit Electronic DB	1 - 3	
Learn Training	2 - 3	
Exploit Reflect	2 - 3	
Connect Educate	1 - 3	
Decide Communities	1 - 3	
Learn Mentor	2 - 3	
Learn Venue	0 - 2	
Connect Evaluation	1 - 2	
Create Suggest	1 - 2	
OR Integer	58 - 67	
Connect Relations	3 - 4	
Learn Capture	1 - 2	
Create Employees	1 - 3	
Link Leadership	0 - 2	
Link Actively	1 - 3	
Create Experiment	2	
Link Monitor	1 - 2	
Decide Co Op	4	
Connect Beliefs	3	
Exploit Simulate	1 - 2	
Link Designated	0 - 2	
Connect Confident	2 - 3	
Connect Resources	4 - 5	
Link Relationship	3	
Create Gap Fix	4	
Learn Offsite	3	
Connect Empower	2 - 3	
Learn Priority	2 - 3	
Create Gap Satisy	1	
Decide Condition	2 - 3	
Connect Buying	0 - 2	
Decide Professional	2 - 3	
Decide Standards	1 - 3	
Learn Priority	4 - 3	
Decide Alliances	0 - 2	

Fig. 6.3.32: Characteristics of Cluster 6 of the Cluster_Model1

The diagram in Fig. 6.3.33 illustrates a cluster analysis of Cluster 4 versus the remaining clusters to determine which variables differentiate, and to what probable extent, the chosen cluster from the rest. This view can identify what is especially important about Cluster 4. From Fig. 6.3.33, it can be seen that, with a probability of around 82%, the 'Disagree' (value = 2) reply to the 'Decide Boundaries' question will be found in Cluster 4. Also, it appears that, with a probability of about 80%, the 'N/A' replies to the 'Decide Alliances' will also be found in Cluster 4. On the other hand, any answer other than 'N/A' to the Decide Alliances question will not be, with 100% probability, in Cluster 4. Additional distinguishing variables can also be found in the figure.



Cluster 1: Cluster 4

Cluster 2: Complement of Clus

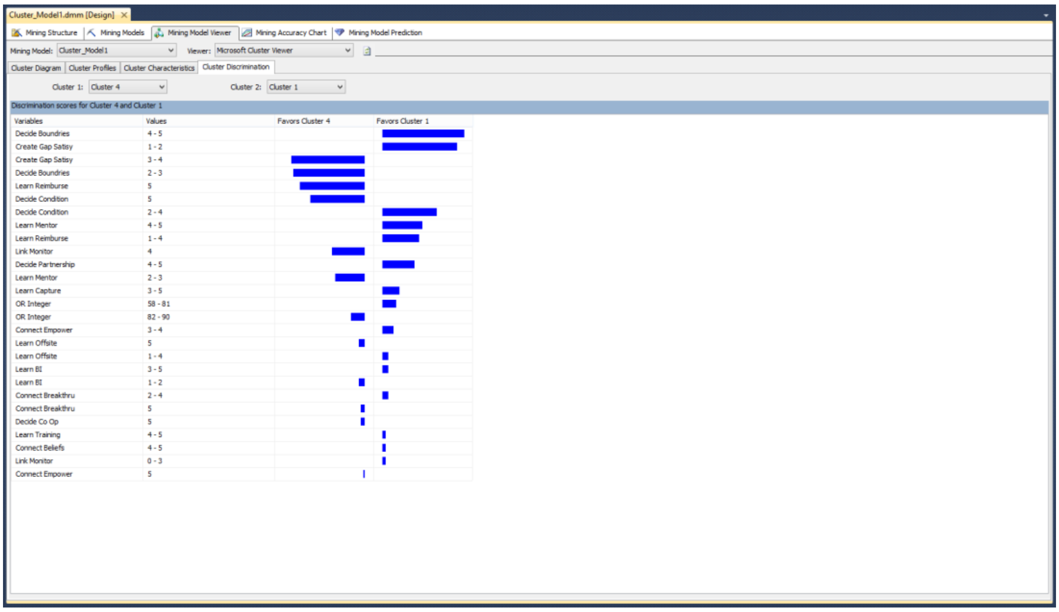
Discrimination scores for Cluster 4 and Complement of Cluster 4

Variables	Values	Favors Cluster 4	Favors Complement of Cluster 4
Decide Alliances	1 - 5		▬
Create Gap Satisy	1 - 4		▬
Decide Partnership	3 - 5		▬
Decide Boundries	3 - 5		▬
Decide Boundries	2	▬	
Decide Alliances	0	▬	
Exploit Consult	2 - 5		▬
Connect Breakthru	2 - 5		▬
Connect Beliefs	3 - 5		▬
Learn Capture	1 - 5		▬
Connect Annual	2 - 5		▬
Connect Empower	2 - 5		▬
Connect Activities	1 - 5		▬
OR Integer	58 - 90		▬
Learn BI	1 - 5		▬
Decide Communities	1 - 5		▬
Decide Chambers	2 - 5		▬
Learn Training	2 - 5		▬
Connect Educate	1 - 5		▬
Learn Mentor	2 - 5		▬
Decide Standards	1 - 5		▬
Learn Venue	0 - 5		▬

Fig. 6.3.33: Difference of specific cluster (Cluster 4) from the general population

As stated by MacLennan (2009, pg. 309), it is very important to contrast the cluster of interest, in this case Cluster 4, with the cluster linked by the strongest link (to Cluster 1) in order to refine the view of the chosen cluster (Cluster 4) (RQ #3). Fig. 6.3.34 illustrates this contrast. High-valued responses to the 'Decide Boundaries' question and low-valued responses to the 'Create Gap Satisfy' question appear to primarily distinguish the two related clusters, with other variables and their probabilities also displayed (the probability percentage is only visible once the cursor is placed over the graph area). Because Clusters 1, and 6 are the clusters holding the entries with the lowest 'OR Integer' scores, and Cluster 4 holds the entries with the highest scores, the strong relationship between Cluster 1 and Cluster 4 is of special interest in terms of contrasting them (RQ #3).

In addition to the comparison of Cluster 4 with Cluster 1, a comparison of Cluster 4, a 'very high score cluster,' with the 'very low score' Cluster 6 provides a fuller picture of which variables and which values favor Cluster 4 (RQ #3). Fig. 6.3.35 illustrates that comparison. It can be seen that the variable 'Exploit Electronic DB' with values 1-3 heavily (with 100% probability) favors Cluster 6, whereas values 4-5 favor, with 100% probability, favor Cluster 4.



Mining Model: Cluster_Model1 Viewer: Microsoft Cluster Viewer

Cluster 1: Cluster 4 Cluster 2: Cluster 1

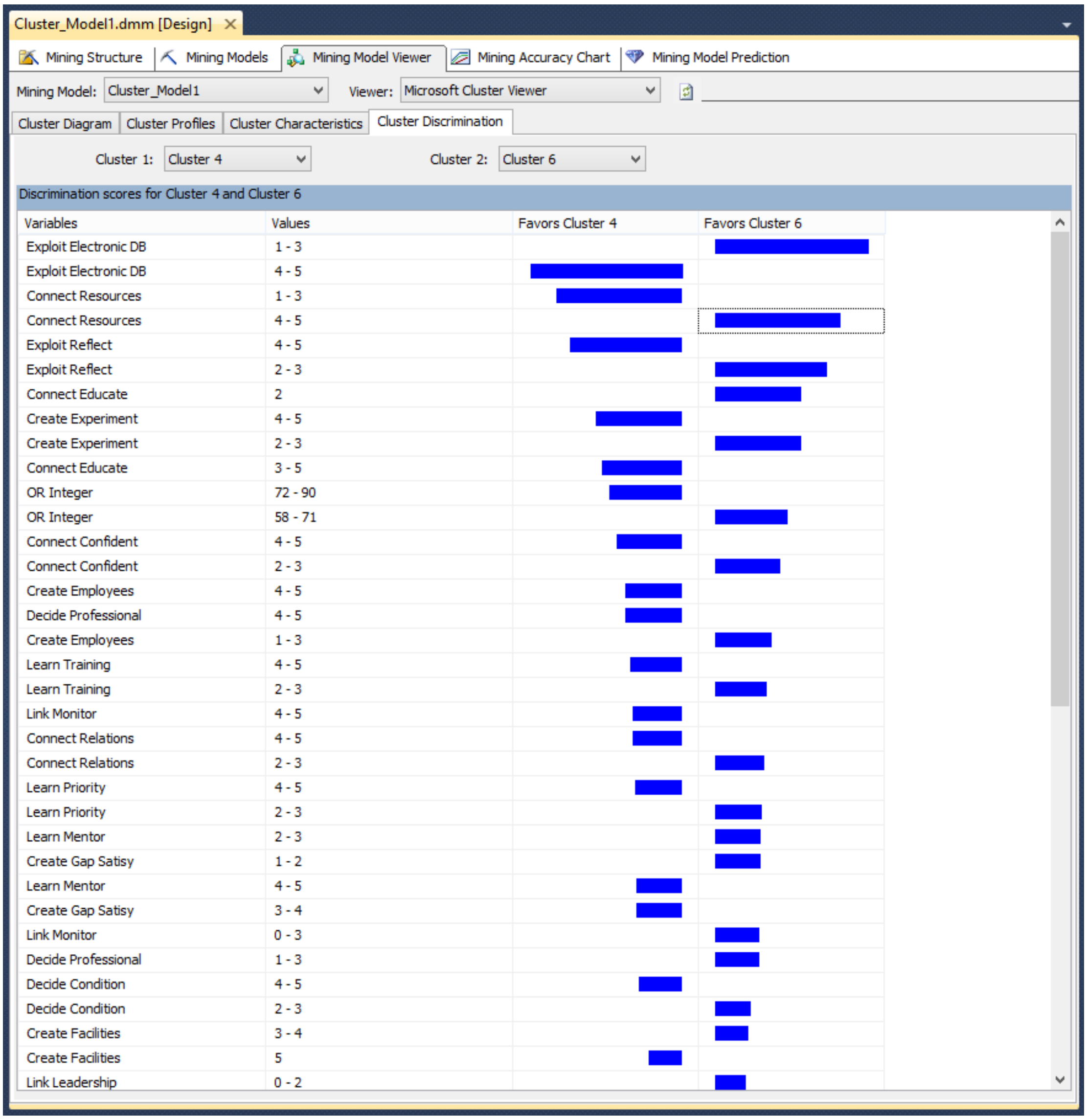
Discrimination scores for Cluster 4 and Cluster 1

Variables	Values	Favors Cluster 4	Favors Cluster 1
Decide Boundries	4 - 5		■
Create Gap Satisy	1 - 2		■
Create Gap Satisy	3 - 4	■	
Decide Boundries	2 - 3	■	
Learn Reimburse	5	■	
Decide Condition	5	■	
Decide Condition	2 - 4		■
Learn Mentor	4 - 5		■
Learn Reimburse	1 - 4		■
Link Monitor	4	■	
Decide Partnership	4 - 5		■
Learn Mentor	2 - 3	■	
Learn Capture	3 - 5		■
OR Integer	58 - 81		■
OR Integer	82 - 90	■	
Connect Empower	3 - 4		■
Learn Offsite	5	■	
Learn Offsite	1 - 4		■
Learn BI	3 - 5		■
Learn BI	1 - 2	■	
Connect Breakthru	2 - 4		■
Connect Breakthru	5	■	
Decide Co Op	5	■	
Learn Training	4 - 5		■
Connect Beliefs	4 - 5		■
Link Monitor	0 - 3		■
Connect Empower	5	■	

Fig. 6.3.34: Contrast of Cluster 4 with Custer 1

The situation with the values of the variable 'Connect Resources' appears to be reversed to that of 'Exploit Electronic DB' variable, with about 82% probability of high values being in Cluster 6 and the same probability of low values being in Cluster 4.

Because, in the resulting model, there are two 'low-scored clusters,' Cluster 1 and Cluster 6, it is therefore advisable to contrast these two clusters in order to obtain additional knowledge about their composition. The results of cluster discrimination between Cluster 1 and Cluster 6 are presented in Fig. 6.3.36. As can be seen, low values in the 'Exploit Electronic DB' variable in Cluster 6, with an 84% probability of preferring Cluster 6, are greatly contrasted with high values for the same variable in Cluster 1 (which also has an 82% preference for that cluster). Given this, it may be advisable to search for differentiating factors between specific variables and the general population, as in Fig. 6.3.33 (RQ #3). Such action would have been taken had there been a sufficiently large amount of input data.



Mining Model: Cluster_Model1

Viewer: Microsoft Cluster Viewer

Cluster 1: Cluster 4

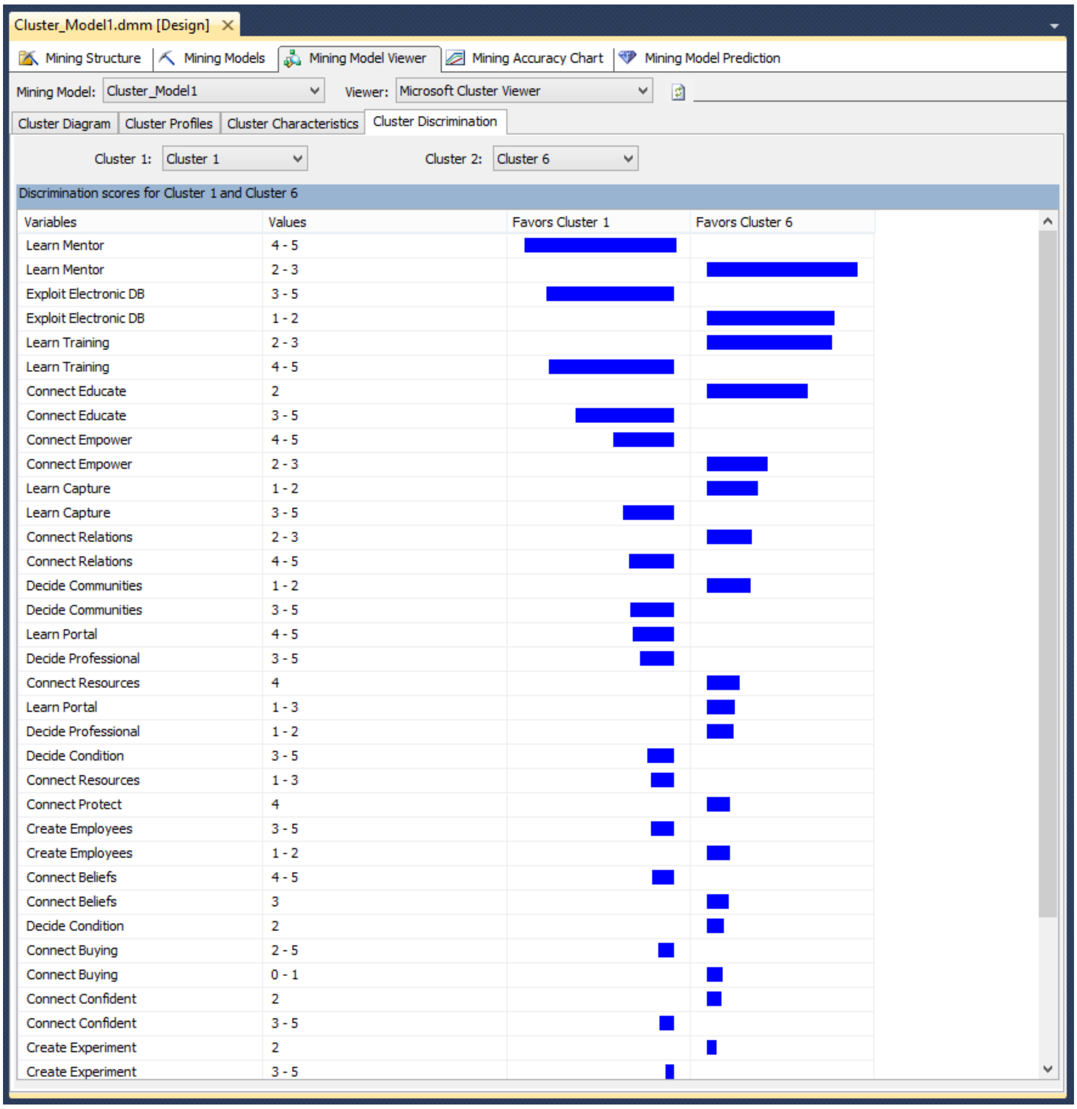
Cluster 2: Cluster 6

Discrimination scores for Cluster 4 and Cluster 6

Variables	Values	Favors Cluster 4	Favors Cluster 6
Exploit Electronic DB	1 - 3		█
Exploit Electronic DB	4 - 5	█	
Connect Resources	1 - 3	█	
Connect Resources	4 - 5		█
Exploit Reflect	4 - 5	█	
Exploit Reflect	2 - 3		█
Connect Educate	2		█
Create Experiment	4 - 5	█	
Create Experiment	2 - 3		█
Connect Educate	3 - 5	█	
OR Integer	72 - 90	█	
OR Integer	58 - 71		█
Connect Confident	4 - 5	█	
Connect Confident	2 - 3		█
Create Employees	4 - 5	█	
Decide Professional	4 - 5	█	
Create Employees	1 - 3		█
Learn Training	4 - 5	█	
Learn Training	2 - 3		█
Link Monitor	4 - 5	█	
Connect Relations	4 - 5	█	
Connect Relations	2 - 3		█
Learn Priority	4 - 5	█	
Learn Priority	2 - 3		█
Learn Mentor	2 - 3		█
Create Gap Satisy	1 - 2		█
Learn Mentor	4 - 5	█	
Create Gap Satisy	3 - 4	█	
Link Monitor	0 - 3		█
Decide Professional	1 - 3		█
Decide Condition	4 - 5	█	
Decide Condition	2 - 3		█
Create Facilities	3 - 4		█
Create Facilities	5	█	
Link Leadership	0 - 2		█

Fig. 6.3.35: Contrast of Cluster 4 with Cluster 6 (contrast between the cluster [4] containing the highest values of ‘OR Integer’ with cluster [6], containing the lowest values)

Note that it is worth pointing out that, as mentioned in Section 6.3.2.2, with the EM method, an element can belong to more than one cluster (clusters can have common elements), so it is advisable to re-construct the model using the K-Means method, where an element is exclusively assigned to one cluster. The differences in the resulting models are presented next. The accuracy of the model, as was the case for the NB-based model, is addressed in Section 5.6.



Mining Model: Cluster_Model1

Viewer: Microsoft Cluster Viewer

Cluster 1: Cluster 1

Cluster 2: Cluster 6

Discrimination scores for Cluster 1 and Cluster 6

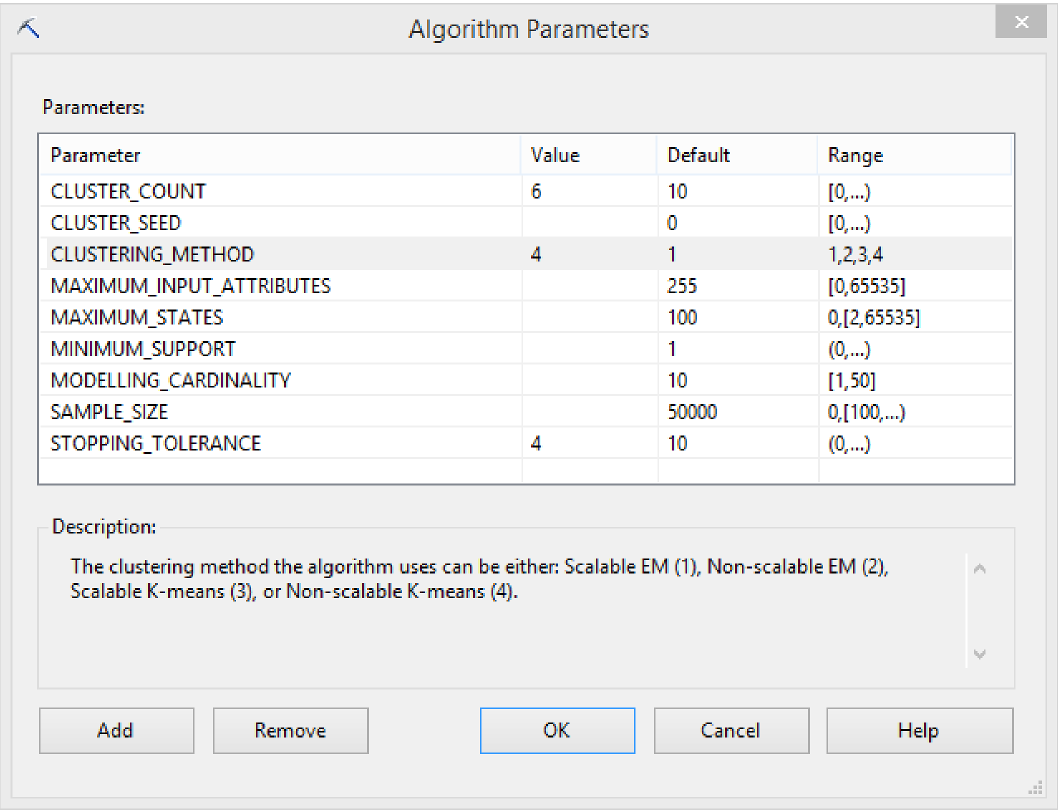
Variables	Values	Favors Cluster 1	Favors Cluster 6
Learn Mentor	4 - 5	■	
Learn Mentor	2 - 3		■
Exploit Electronic DB	3 - 5	■	
Exploit Electronic DB	1 - 2		■
Learn Training	2 - 3		■
Learn Training	4 - 5	■	
Connect Educate	2		■
Connect Educate	3 - 5	■	
Connect Empower	4 - 5	■	
Connect Empower	2 - 3		■
Learn Capture	1 - 2		■
Learn Capture	3 - 5	■	
Connect Relations	2 - 3		■
Connect Relations	4 - 5	■	
Decide Communities	1 - 2		■
Decide Communities	3 - 5	■	
Learn Portal	4 - 5	■	
Decide Professional	3 - 5	■	
Connect Resources	4		■
Learn Portal	1 - 3		■
Decide Professional	1 - 2		■
Decide Condition	3 - 5	■	
Connect Resources	1 - 3	■	
Connect Protect	4		■
Create Employees	3 - 5	■	
Create Employees	1 - 2		■
Connect Beliefs	4 - 5	■	
Connect Beliefs	3		■
Decide Condition	2		■
Connect Buying	2 - 5	■	
Connect Buying	0 - 1		■
Connect Confident	2		■
Connect Confident	3 - 5	■	
Create Experiment	2		■
Create Experiment	3 - 5	■	

Fig. 6.3.36: Contrasting both 'low score' clusters: Cluster 1 with Cluster 6

K-Means Method

With all other parameter values unchanged, the value of the CLUSTERING_METHOD parameter was changed from 2 (non-scalable EM) to 4 (non-scalable K-means). Fig. 6.3.37 illustrates the values of all parameters.

The K-means-based clustering model that was generated segmented the data into five, instead six clusters, as shown in Fig. 6.3.38. The figure displays the cluster diagram, with 'Population' as the shading variable. As can be seen, Cluster 1 has the largest density (34%) of the samples and, considering the 'Population,' it has the strongest link to Cluster 4.



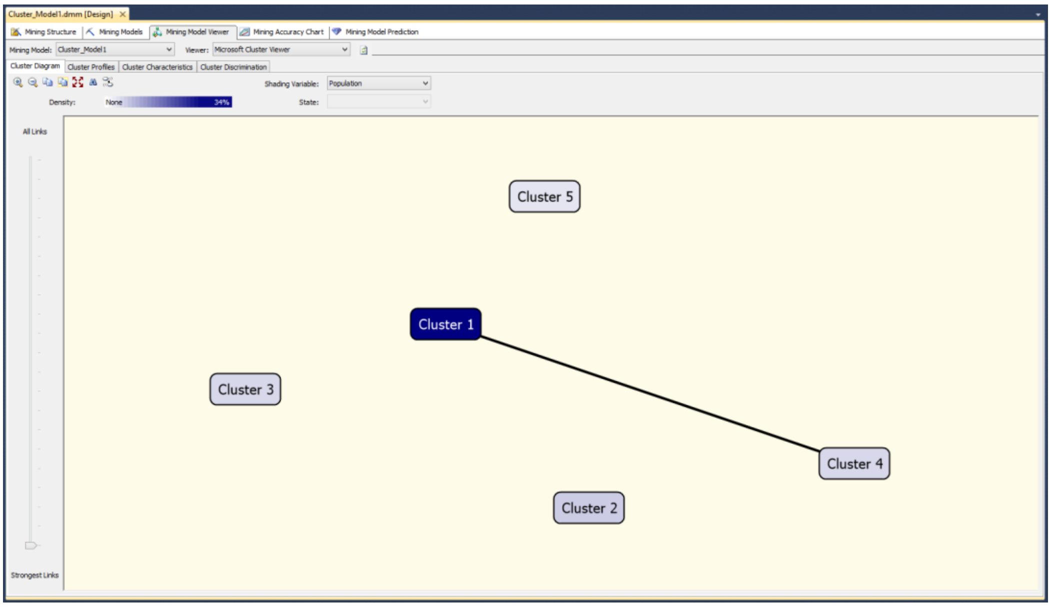
Parameters:

Parameter	Value	Default	Range
CLUSTER_COUNT	6	10	[0,...)
CLUSTER_SEED		0	[0,...)
CLUSTERING_METHOD	4	1	1,2,3,4
MAXIMUM_INPUT_ATTRIBUTES		255	[0,65535]
MAXIMUM_STATES		100	0,[2,65535]
MINIMUM_SUPPORT		1	(0,...)
MODELLING_CARDINALITY		10	[1,50]
SAMPLE_SIZE		50000	0,[100,...)
STOPPING_TOLERANCE	4	10	(0,...)

Description:

The clustering method the algorithm uses can be either: Scalable EM (1), Non-scalable EM (2), Scalable K-means (3), or Non-scalable K-means (4).

Fig. 6.3.37: Values of clustering algorithm with K-means method selected



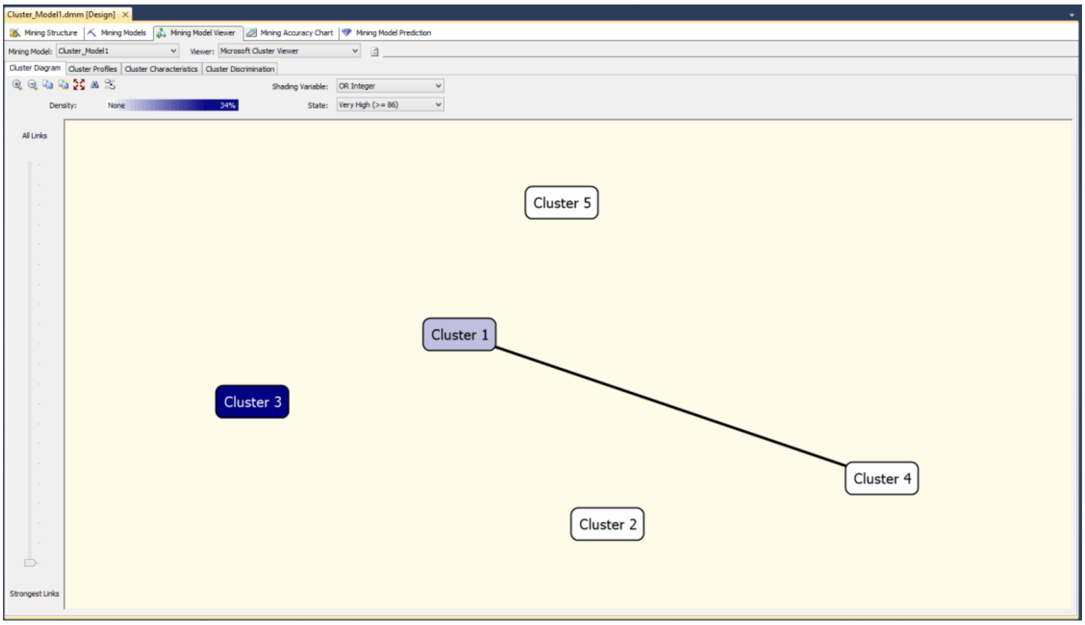
Mining Model: Cluster_Model1

Viewer: Microsoft Cluster Viewer

Shading Variable: Population

Fig. 6.3.38: Cluster diagram – K-means method using shading variable of ‘Population’

Figures 6.3.39 and 6.3.40 illustrate the cluster diagram for the shading variable ‘OR Integer’ with the very high and very low values, respectively. With respect to Fig. 6.3.39, the highest density of the very high values of 'OR Integer' (those greater than or equal to 86) are in Cluster 3, with the strongest link excluding Cluster 3 and being between Cluster 1 (with some concentration of very high values in 'OR Integer' variable) and Cluster 4 (RQ #3). Clusters 2, 4, and 5 appear to have a very limited concentration of very high values in the 'OR Integer' variable.



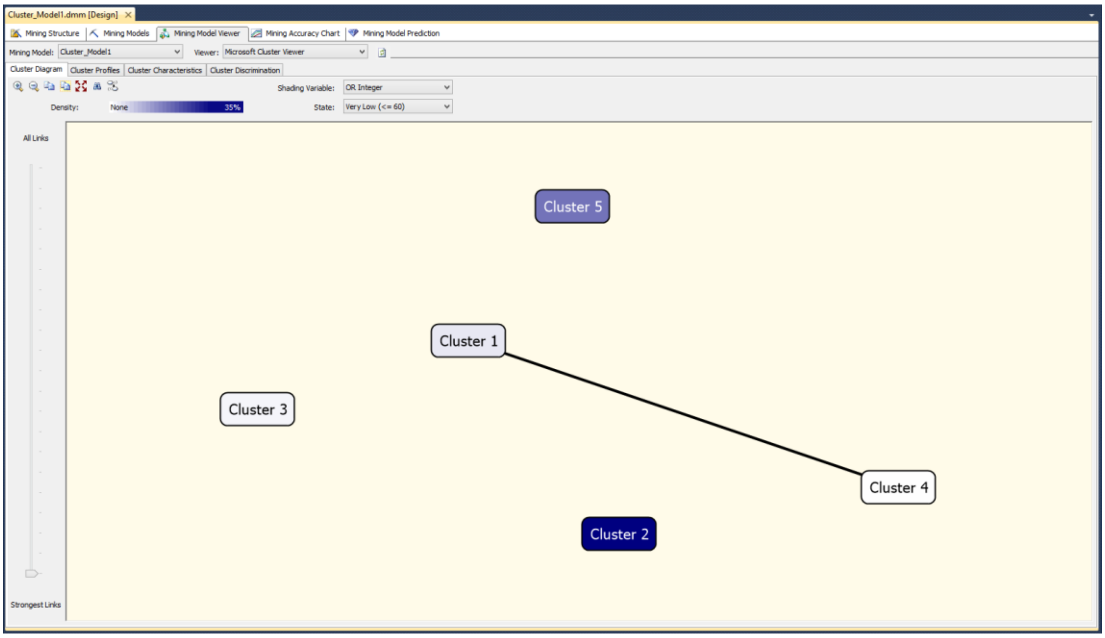
Mining Model: Cluster_Model1
Viewer: Microsoft Cluster Viewer
Shading Variable: OR Integer
State: Very High (>= 86)

Fig. 6.3.39: Cluster diagram for very high values in variable 'OR Integer' with the strongest link shown

The cluster diagram pictured in Fig. 6.3.40 illustrates the density of the 'OR Integer' variable containing very low values (those less than or equal to 60). As can be seen (not very easily, however, in the case of Cluster 1 and 3, due to the light color used), all clusters, with the exception of Cluster 4, contain a very low-valued 'OR Integer'. The strongest relationship shown is, as in the prior figure, between Cluster 1 and Cluster 4. Finally, Cluster 2 is shown as having a density of 35% of very low values in the 'OR Integer' variable – the highest density cluster (RQ #3).

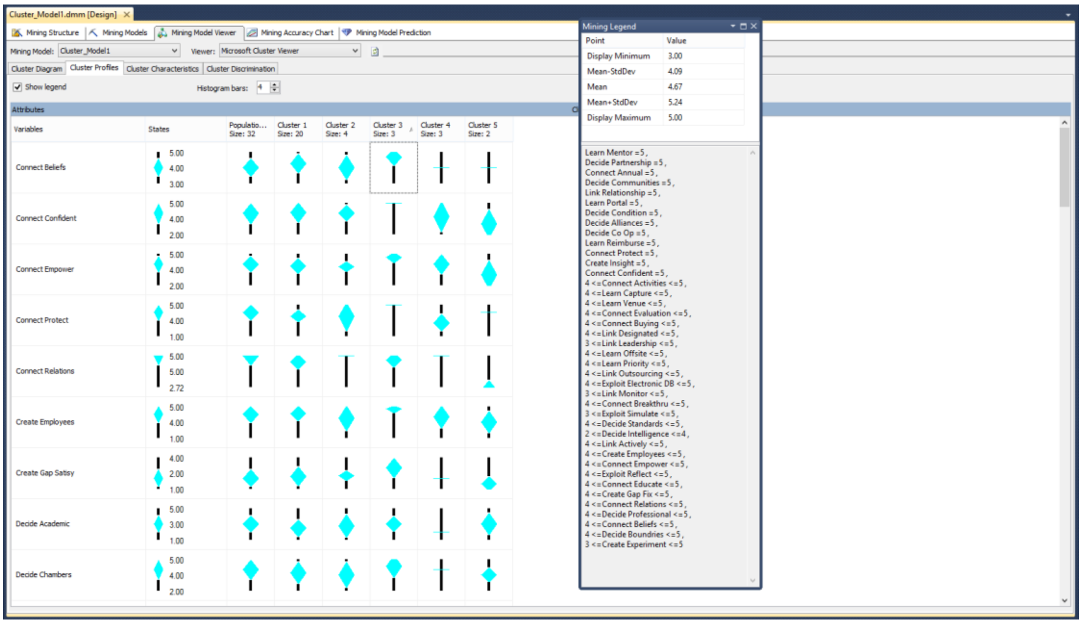
Cluster profiles for a K-means-based model are displayed in Fig. 6.3.41 and are similar to those presented in Fig. 6.3.30. The 'States' column, which has not been discussed previously for the continuous type of variable, provides information about the distribution of the values in a variable across all clusters.

(An example of distribution information displayed for a variable is shown in Appendix VIII, Fig. A8.25.).



Mining Model: Cluster_Model1
Viewer: Microsoft Cluster Viewer
Shading Variable: OR Integer
State: Very Low (<= 60)

Fig. 6.3.40: Cluster diagram for very low values in variable ‘OR Integer’, with the strongest link shown



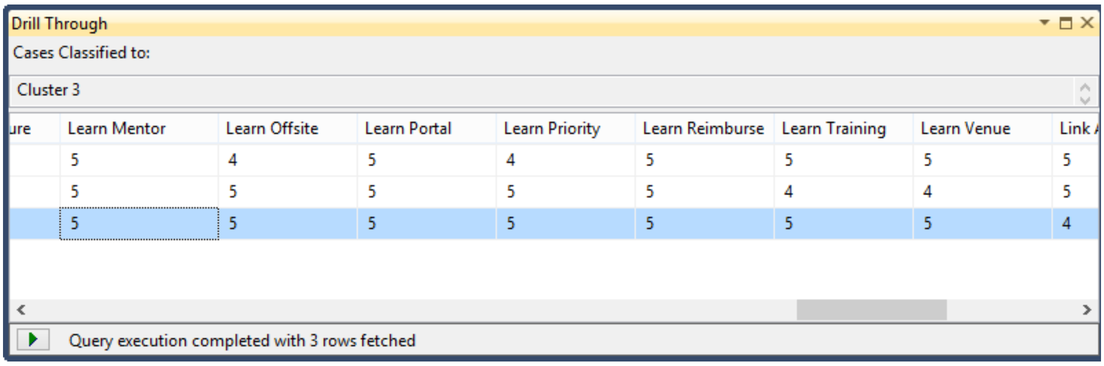
Mining Model: Cluster_Model1
Viewer: Microsoft Cluster Viewer
☑ Show legend
Histogram bars: 4

Point	Value
Display Minimum	3.00
Mean-StdDev	4.09
Mean	4.67
Mean+StdDev	5.24
Display Maximum	5.00

Learn Mentor =5,
Decide Partnership =5,
Connect Annual =5,
Decide Communities =5,
Link Relationship =5,
Learn Portal =5,
Decide Condition =5,
Decide Alliances =5,
Decide Co Op =5,
Learn Reimburse =5,
Connect Protect =5,
Create Insight =5,
Connect Confident =5,
4 <=Connect Activities <=5,
4 <=Learn Capture <=5,
4 <=Learn Venue <=5,
4 <=Connect Evaluation <=5,
4 <=Connect Buying <=5,
4 <=Link Designated <=5,
3 <=Link Leadership <=5,
4 <=Learn Offsite <=5,
4 <=Learn Priority <=5,
4 <=Link Outsourcing <=5,
4 <=Exploit Electronic DB <=5,
3 <=Link Monitor <=5,
4 <=Connect Breakthru <=5,
3 <=Exploit Simulate <=5,
4 <=Decide Standards <=5,
2 <=Decide Intelligence <=4,
4 <=Link Actively <=5,
4 <=Create Employees <=5,
4 <=Connect Empower <=5,
4 <=Exploit Reflect <=5,
4 <=Connect Educate <=5,
4 <=Create Gap Fix <=5,
4 <=Connect Relations <=5,
4 <=Decide Professional <=5,
4 <=Connect Beliefs <=5,
4 <=Decide Boundries <=5,
3 <=Create Experiment <=5

Fig. 6.3.41: Cluster profiles with mining legend

The mining legend pictured in Fig. 6.3.41, to the right of the cluster profiles, shows the values of variables (in decreasing order of values) that make up a given cluster (Cluster 3), in addition to the basic statistic about the selected variable-cluster pair ('Connect Beliefs'-Cluster 3).


Drill Through

Cases Classified to:

Cluster 3

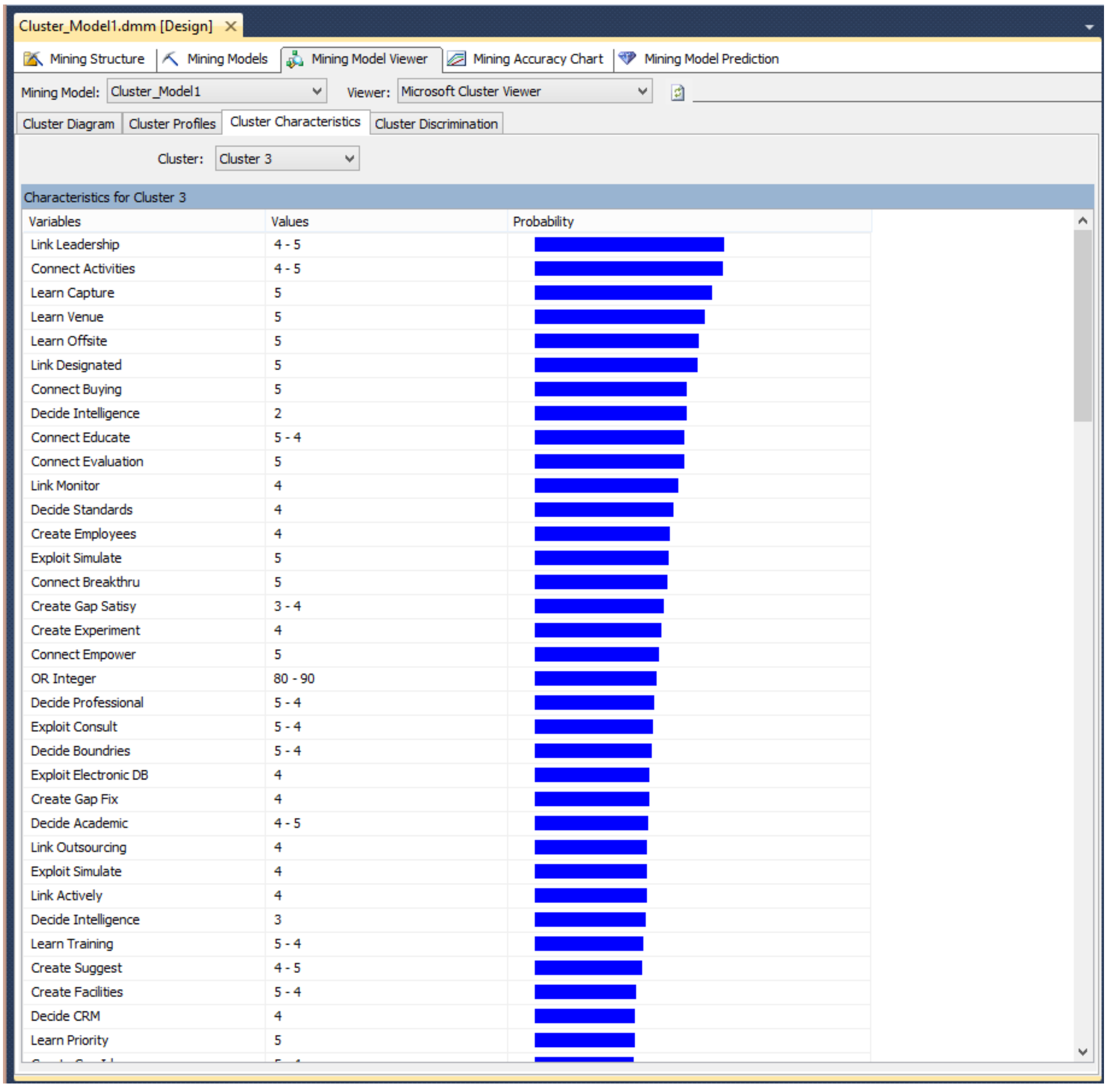
ure	Learn Mentor	Learn Offsite	Learn Portal	Learn Priority	Learn Reimburse	Learn Training	Learn Venue	Link A
	5	4	5	4	5	5	5	5
	5	5	5	5	5	4	4	5
	5	5	5	5	5	5	5	4

Query execution completed with 3 rows fetched

Fig. 6.3.42: Cluster 3 contents using the 'Drill Through' option

Fig. 6.3.43, below, is similar to Fig. 6.3.31 in that it presents the characteristics of the cluster (in this case, Cluster 3) with the highest number of high scores in the 'OR Integer' variable. While there is limited value in comparing the two displays due to the very low amount of input data (RQ #5) used in the model-building process, it can be noted that the only two variables from Cluster 4 (in the model based on non-scalable EM) that were listed in somewhat 'top positions' in Cluster 3 (of the K-means based model) were 'Link Monitor' and 'Create Experiment'. (Section 6.2.6, among other topics, discusses the issue of repeatability of outcomes between models.)

Cluster discrimination between the cluster with the highest 'OR Integer' values, Cluster 3, and all other clusters is presented in Fig. 6.3.44. While the 'OR Integer' variable itself is of little interest in the figure, the 'Create Gap Satisfy' variable favors 100% presence of values 1-4, but not, however, the value of 5, in all clusters other than Cluster 3. The variable 'Decide Chambers,' with a probability of about 83%, favors other clusters for values 2-5. The remaining two variables favor other clusters with about 80% probability with the variable 'Create Suggest' across all values, with the exception of the value zero (for the 'N/A' answer).



Mining Model: Cluster_Model1

Viewer: Microsoft Cluster Viewer

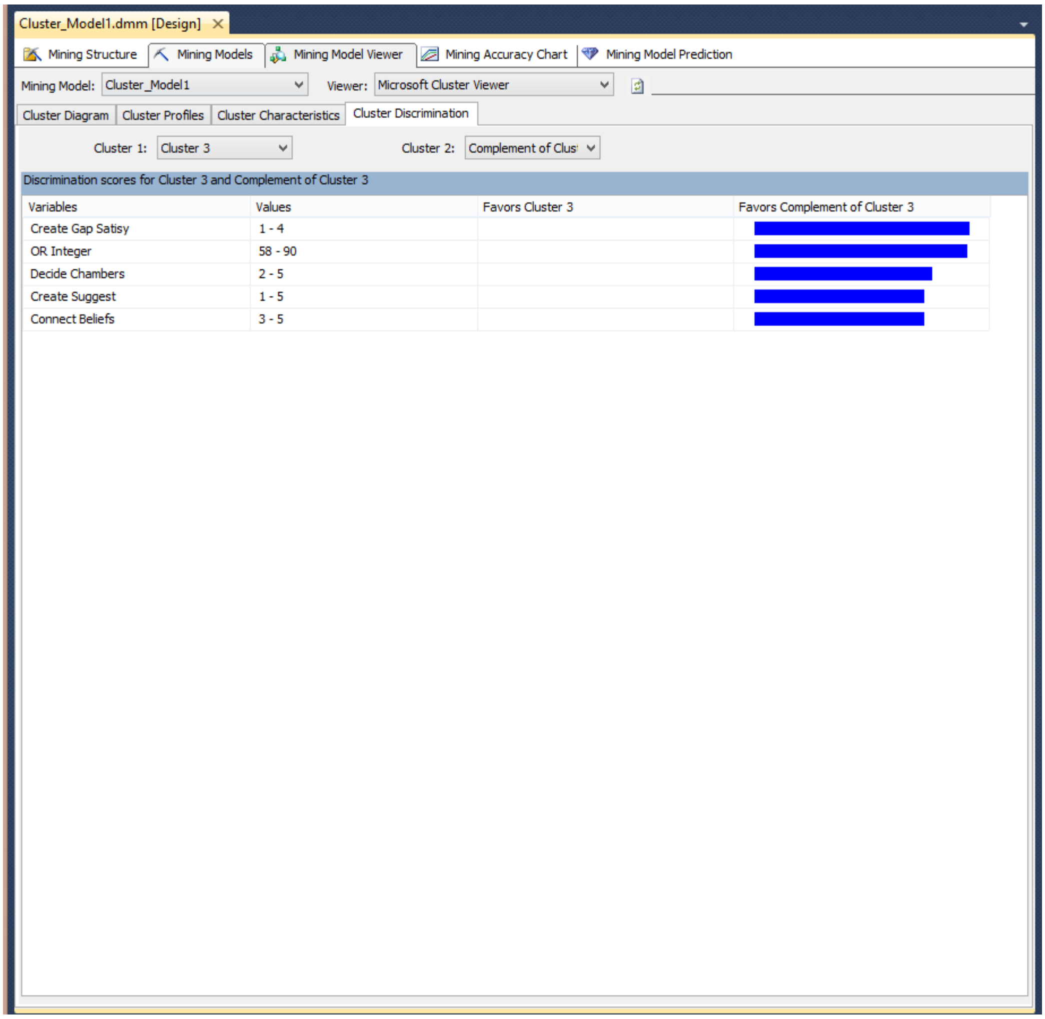
Cluster: Cluster 3

Characteristics for Cluster 3

Variables	Values	Probability
Link Leadership	4 - 5	
Connect Activities	4 - 5	
Learn Capture	5	
Learn Venue	5	
Learn Offsite	5	
Link Designated	5	
Connect Buying	5	
Decide Intelligence	2	
Connect Educate	5 - 4	
Connect Evaluation	5	
Link Monitor	4	
Decide Standards	4	
Create Employees	4	
Exploit Simulate	5	
Connect Breakthru	5	
Create Gap Satisy	3 - 4	
Create Experiment	4	
Connect Empower	5	
OR Integer	80 - 90	
Decide Professional	5 - 4	
Exploit Consult	5 - 4	
Decide Boundries	5 - 4	
Exploit Electronic DB	4	
Create Gap Fix	4	
Decide Academic	4 - 5	
Link Outsourcing	4	
Exploit Simulate	4	
Link Actively	4	
Decide Intelligence	3	
Learn Training	5 - 4	
Create Suggest	4 - 5	
Create Facilities	5 - 4	
Decide CRM	4	
Learn Priority	5	

Fig. 6.3.43: Cluster 3 characteristics

The discrimination between the highest valued cluster, in terms of answer points, (Cluster 3) and the lowest valued cluster (Cluster 2) does not introduce any unexpected results, at least in the first dozen or so positions (RQ #3). The variables favoring Cluster 3 all have the highest values and those in Cluster 2, generally, the lowest. The most influential variables, however, differ entirely from the variables compared in the model that used the non-scalable EM algorithm (RQ #5).



Mining Model: Cluster_Model1

Viewer: Microsoft Cluster Viewer

Cluster 1: Cluster 3

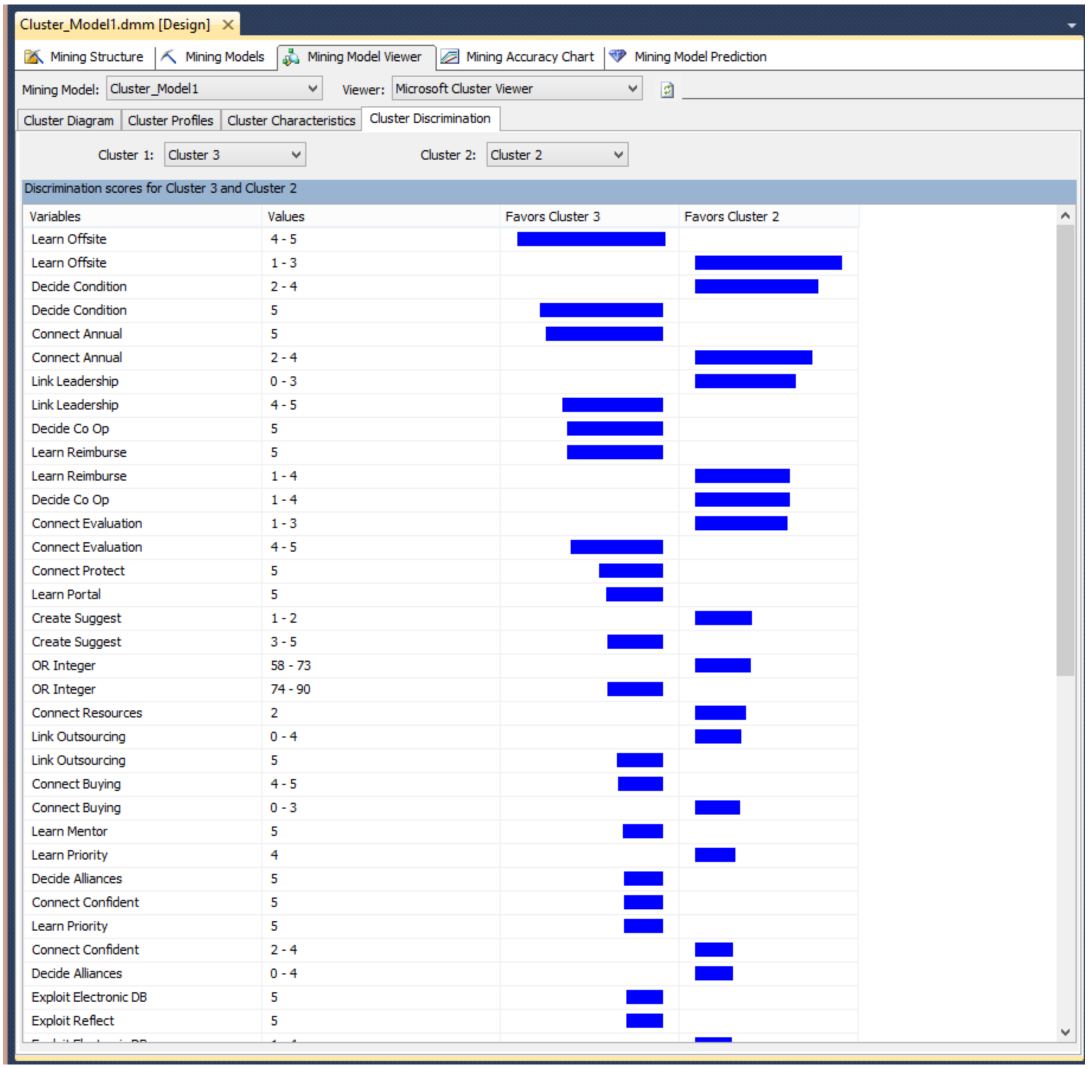
Cluster 2: Complement of Clus

Discrimination scores for Cluster 3 and Complement of Cluster 3

Variables	Values	Favors Cluster 3	Favors Complement of Cluster 3
Create Gap Satisy	1 - 4		▬
OR Integer	58 - 90		▬
Decide Chambers	2 - 5		▬
Create Suggest	1 - 5		▬
Connect Beliefs	3 - 5		▬

Fig. 6.3.44: Cluster discrimination: Cluster 3 versus all other clusters

While not presented in this research, contrasting the composition between clusters using the cluster discrimination can also be beneficial when comparing, for example, the ‘highest valued’ cluster with just one ‘slightly worse’ cluster. Such a comparison can highlight what is missing, in terms of KM processes, in an organization that is attempting to achieve greater levels of resilience (RQ #3). Similarly, on the other side of the spectrum, one can inspect the two ‘lowest valued’ clusters to determine the factors (in this case, KM processes) that, when not addressed, can further reduce an organization’s resilience.

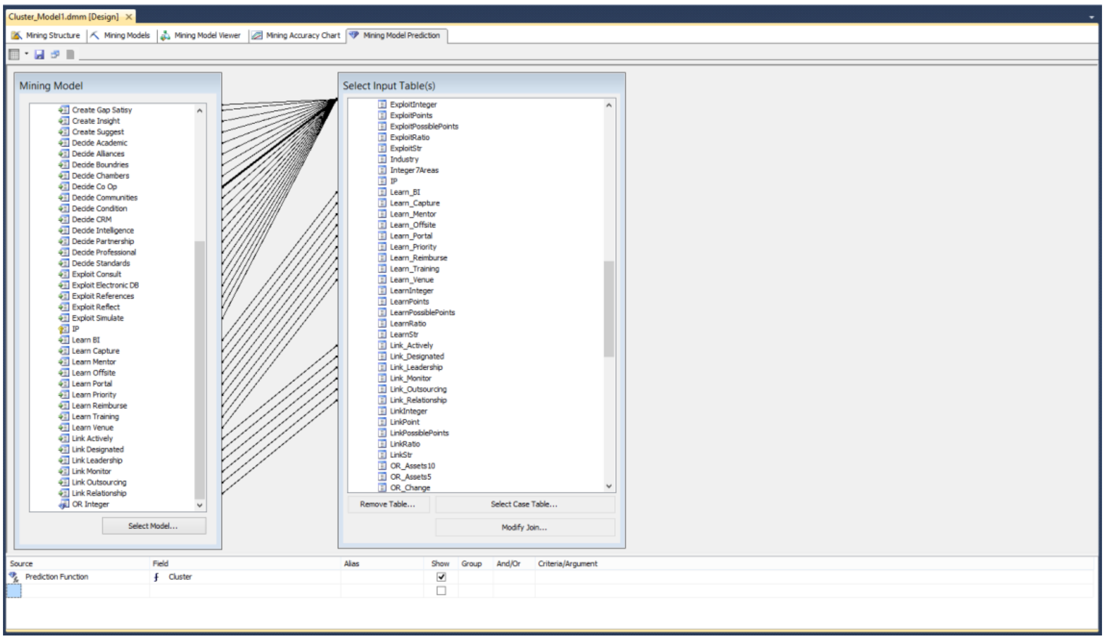


Discrimination scores for Cluster 3 and Cluster 2

Variables	Values	Favors Cluster 3	Favors Cluster 2
Learn Offsite	4 - 5	■	
Learn Offsite	1 - 3		■
Decide Condition	2 - 4		■
Decide Condition	5	■	
Connect Annual	5	■	
Connect Annual	2 - 4		■
Link Leadership	0 - 3		■
Link Leadership	4 - 5	■	
Decide Co Op	5	■	
Learn Reimburse	5	■	
Learn Reimburse	1 - 4		■
Decide Co Op	1 - 4		■
Connect Evaluation	1 - 3		■
Connect Evaluation	4 - 5	■	
Connect Protect	5	■	
Learn Portal	5	■	
Create Suggest	1 - 2		■
Create Suggest	3 - 5	■	
OR Integer	58 - 73		■
OR Integer	74 - 90	■	
Connect Resources	2		■
Link Outsourcing	0 - 4		■
Link Outsourcing	5	■	
Connect Buying	4 - 5	■	
Connect Buying	0 - 3		■
Learn Mentor	5	■	
Learn Priority	4		■
Decide Alliances	5	■	
Connect Confident	5	■	
Learn Priority	5	■	
Connect Confident	2 - 4		■
Decide Alliances	0 - 4		■
Exploit Electronic DB	5	■	
Exploit Reflect	5	■	

Fig. 6.3.45: Cluster discrimination: Cluster 3 versus Cluster 2

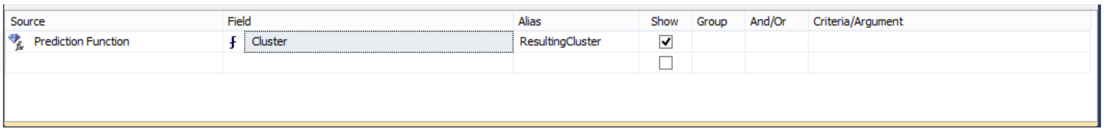
For predictive purposes, the same row of data used for NB_Model2 and all other models was used as an input. (The input data for predictive purposes was discussed when presenting Fig. 6.3.23. In addition, the model’s source construct is presented in Appendix VIII, Fig. A 8.15.) For the purpose of this prediction, individual answers stored in the input table (tbl_NBModel2_Predict) were mapped to the mining model, as shown in Fig. 6.3.46.



Source	Field	Alias	Show	Group	And/Or	Criteria/Argument
Prediction Function	Cluster		☑			
			□			

Fig. 6.3.46: Illustration of the mining model prediction using Cluster_Model1: mapping of the input fields to the model's fields

The prediction function and the results of the application of the 'Cluster' prediction function are presented in figures 6.3.47 and 6.3.48. The data present in the input table was determined by the algorithm to belong to Cluster 1. (The query generated by SQL Server in order to arrive at the prediction is presented in Appendix VIII, Fig. A8.26).

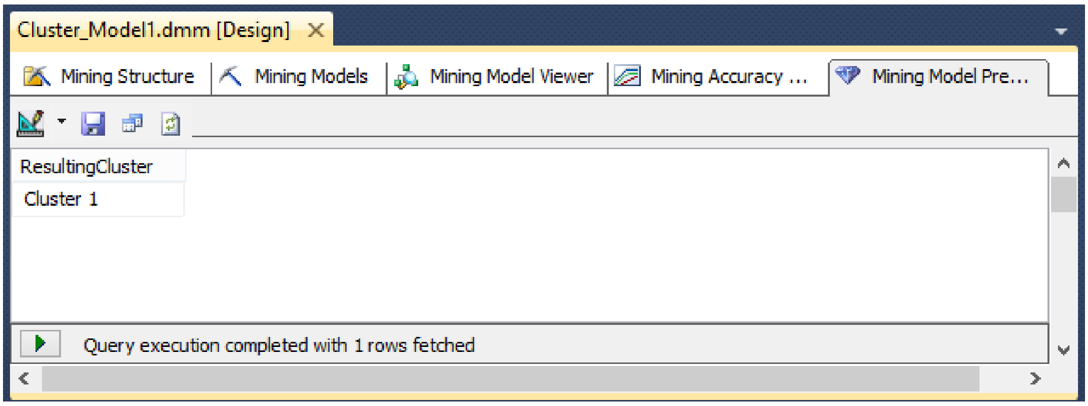


Source	Field	Alias	Show	Group	And/Or	Criteria/Argument
Prediction Function	Cluster	ResultingCluster	☑			
			□			

Fig. 6.3.47: Illustration of the use of the 'Cluster' function

Classification of the answers into a cluster resulted in the outcome presented in Fig. 6.3.48.

While the limited amount of data in the model, and therefore the lower quality of the model discussed in Chapter 5.6, makes the analysis of the actual results of limitted value (RQ #5), it worth pointing out that the results of the prediction can be compared, among other things, with the cluster characteristics of the resultant cluster (Cluster 1). Fig. 6.3.49 shows some of the top-most characteristics of Cluster 1.



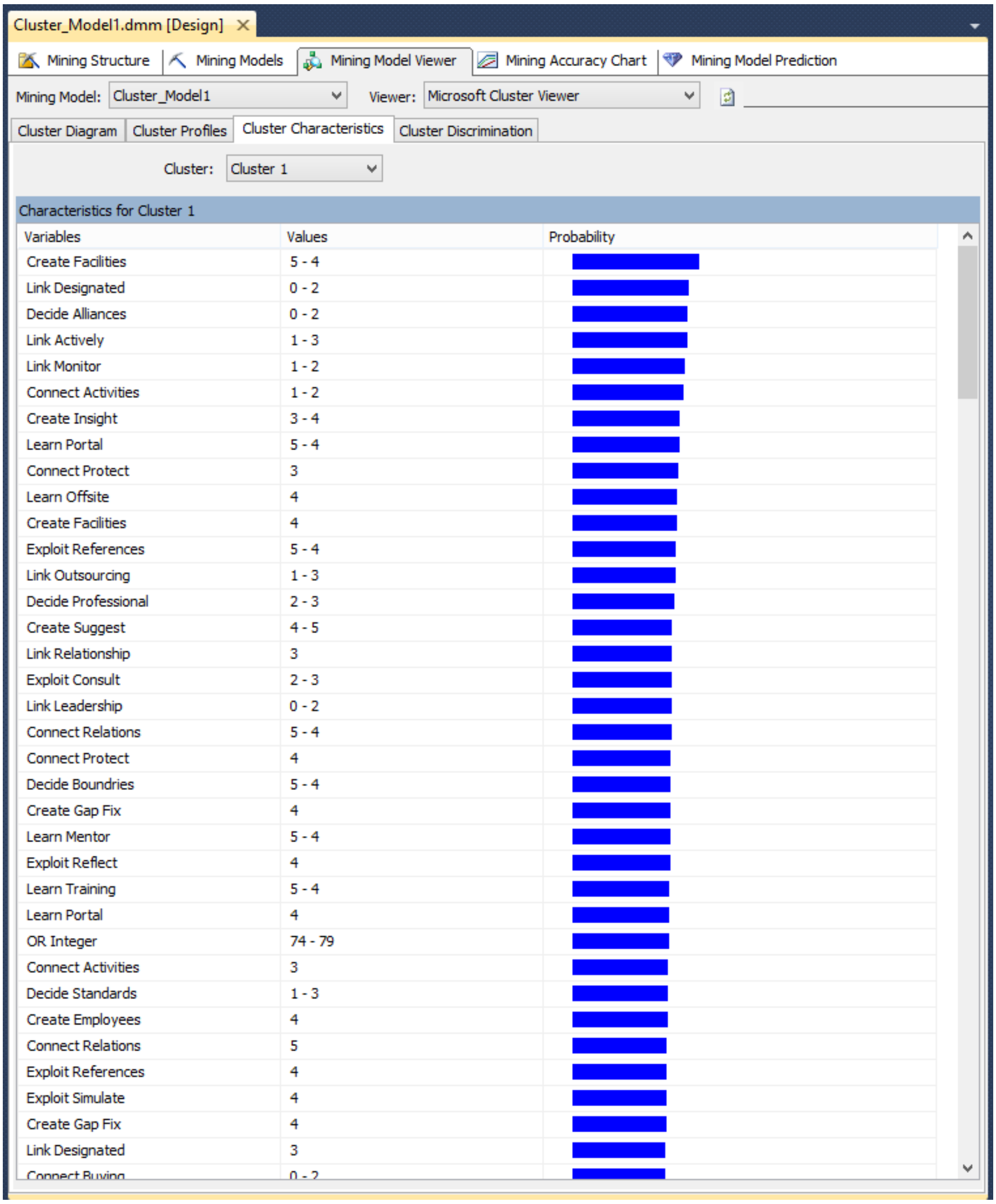
ResultingCluster
Cluster 1

Query execution completed with 1 rows fetched

Fig. 6.3.48: The result of the ‘Cluster’ function

Even with a sufficient amount of data for accurate model construction, the fact that that every value of input data will match the value of the variable of the predicted cluster means that the use of the ‘Predict Only’ variable (in this case the variable ‘OR Integer’) can help to provide insight into the prediction. In the case of the value of the ‘OR Integer’ for Cluster 1 (Fig. 6.3.49), the range is between 74 and 79. (The link between the ‘OR Integer’ in the input table and the matching field in the model has been removed, so the field plays no role in prediction.)

With regard to determining the ‘OR Integer’ value using SQL Server’s clustering algorithm, the following sources for the prediction were set up (Fig. 6.3.50). As listed in Fig. 6.3.50, the ‘Cluster_Model1’ source uses the clustering model and the data present there to arrive at the ‘OR Integer’ value (labeled as ‘Model Data OR Score’), and the ‘tbl_NBModle2_Predict’ (the input table containing single data row) uses the data in the input table to arrive at the ‘OR Integer’ value (labeled as ‘Input Table Data OR Score’).


Cluster_Model1.dmm [Design]

Mining Structure | Mining Models | Mining Model Viewer | Mining Accuracy Chart | Mining Model Prediction

Mining Model: Cluster_Model1 Viewer: Microsoft Cluster Viewer

Cluster Diagram | Cluster Profiles | Cluster Characteristics | Cluster Discrimination

Cluster: Cluster 1

Characteristics for Cluster 1

Variables	Values	Probability
Create Facilities	5 - 4	
Link Designated	0 - 2	
Decide Alliances	0 - 2	
Link Actively	1 - 3	
Link Monitor	1 - 2	
Connect Activities	1 - 2	
Create Insight	3 - 4	
Learn Portal	5 - 4	
Connect Protect	3	
Learn Offsite	4	
Create Facilities	4	
Exploit References	5 - 4	
Link Outsourcing	1 - 3	
Decide Professional	2 - 3	
Create Suggest	4 - 5	
Link Relationship	3	
Exploit Consult	2 - 3	
Link Leadership	0 - 2	
Connect Relations	5 - 4	
Connect Protect	4	
Decide Boundries	5 - 4	
Create Gap Fix	4	
Learn Mentor	5 - 4	
Exploit Reflect	4	
Learn Training	5 - 4	
Learn Portal	4	
OR Integer	74 - 79	
Connect Activities	3	
Decide Standards	1 - 3	
Create Employees	4	
Connect Relations	5	
Exploit References	4	
Exploit Simulate	4	
Create Gap Fix	4	
Link Designated	3	
Connect Buying	0 - 2	

Fig. 6.3.49: Some of the characteristics of Cluster 1 (Cluster_Model1)

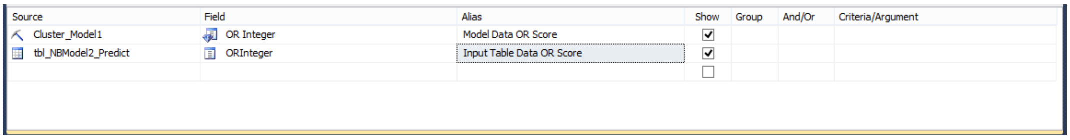
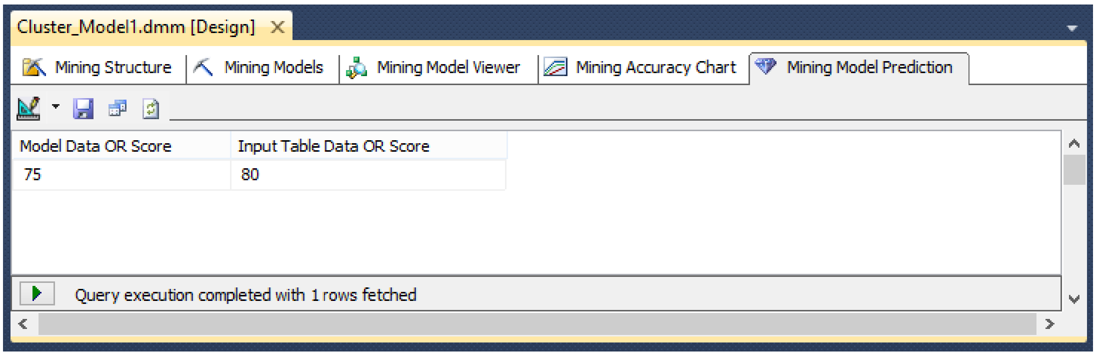

Source	Field	Alias	Show	Group	And/Or	Criteria/Argument
Cluster_Model1	OR Integer	Model Data OR Score	☑			
tbl_NBModel2_Predict	ORInteger	Input Table Data OR Score	☑			
			□			

Fig. 6.3.50: Sources for predicting the value of the ‘OR Integer’ variable

The results of the prediction are shown in Fig. 6.3.51 and Appendix VIII; Fig. A8.27 contains the query used to arrive with the results. Consistent with the ‘OR Integer’ value of Cluster 1, the model, given very limited data, returned the value of 75, and the value of ‘OR Integer’ in the input table was determined by the prediction to be 80. (That is, there was only a difference of 5 between the predicted value of 75 and the actual OR Score of 80.)



Model Data OR Score	Input Table Data OR Score
75	80

Query execution completed with 1 rows fetched

Fig. 6.3.51: Results of predicting ‘OR Integer’ value

6.3.3 Individual KM Processes

Two additional aspects that need to be discussed with respect to this research question are the grouping of KM activities (with a KM activity, for the purpose of this discussion, being an activity addressed by a single question in the questionnaire) into alternative to the competence areas grouping and the use of individual KM activities in modeling.

It needs to be pointed out that the construction of the DM models did not require the use the McKenzie and van Winkelen (2004) framework or the KM process model proposed by Burnett (2004, pg. 29). Instead, one could rely on the DM clustering algorithm to segment the questionnaire replies into clusters where the similarity of the questions in a cluster is maximized and similarity outside of the cluster in minimized.

To illustrate the use of DM clustering technique, the Cluster_Model1 is constructed in two ways: the first uses six output clusters, to match the number of competence areas, and the second way allows the algorithm to arrive at the optimal number of output clusters. The model, its construction and the composition of groupings are described in Section 6.3.2. There were a number of attempts at model creation with predetermined number of clusters, as the algorithm parameters proved to be an obstacle that had to be overcome (RQ #5). Once successfully constructed, the inspection of the composition of each of the resultant cluster was possible; the inspection was performed using the ‘Cluster

Profile,', 'Cluster Characteristics' and 'Cluster Discrimination' functions of MS SQL Server. Because of Microsoft's specific implementation of these functions, selection of the OR output variable as 'Predict Only' allowed for the inspection of results in the context of the output variable 'OR Integer' – meaning that 'low scoring' and 'high scoring' clusters were identified, inspected and later compared against each other. In Cluster_Model1, it was found that Cluster 4 had the highest concentration of the 'high-scoring' (4 or 5 points) answers, and Cluster 6 had the highest concentration of the 'low-scoring' (1-3) answers. As an example of the possibility of retrieving information, it has been shown that organizations that do not exploit electronic databases for the recording and retrieval of the lessons learned have an 87% chance of being placed in Cluster 6; organizations that do not offer regular employee training have an 85% chance. Similar to the other DM models discussed in this research, clustering allows identifying the probability of a given KM activity favoring a specific cluster (RQ #3). While the discrimination of the cluster with the highest concentration of high scores with the cluster with the highest concentration of low scores makes natural sense, the comparison between other clusters can point out the differences between KM activities that result in them 'being placed in a higher cluster', meaning that an organization is more resilient (RQ #3). Figures 6.3.28 – 6.3.36 and 6.3.39 – Fig. 6.3.45 in Section 6.3.2 further illustrate this discussion.

The prediction functionality of the clustering algorithm serves two purposes in this research. First, given a set of replies to the questionnaire's questions and using the clustering algorithm's predictability functionality, it is possible to predict into which cluster the responding organization will fall. Secondly, using the responses regarding KM activities, it is possible to predict an organization's 'OR Score'. For the data set common to all predictions, the'OR Score' predicted by the MCluster_Model1 was 75, while the 'OR Score' manually computed for the data set was 80.

It has been found in this research that the DM clustering method is a highly effective approach for segmenting input data into related groups.

While Cluster_Model1 worked well (as expected, due to the nature of the algorithm) with individual KM activities, the same cannot be said of the other algorithms used in this research (namely Naïve Bayes, neural networks and decision trees) (RQ #5). Some of the issues associated with the management of a large number of input variables are presented in Appendix VIII and figures A8.15 to A8.20 and A8.22. The bigger problem, which perhaps is magnified as a result of the small input data set used in the research (as, typically the larger the number of variables, the lager input data set is needed), relates to the interpretation of the results. As shown in A8.47, examining a network diagram that contains 52 variables is difficult (RQ #5). Similar difficulties in interpretation would be encountered when attempting to inspect the composition of the 52 clusters, for example. For this reason, it was decided to group input variables into competence areas for the purpose of this research. Should there be enough input data available, the analysis of individual KM activities could be considered using all of, or a subset of, the responses, such as the subset of responses contained within the competence area with the highest 'OR Score'.

6.4 Data Mining: Neural Network (NN)

Inspired by the inner workings of the human brain, the neural network algorithm was developed in the 1960s 'to imitate a type of a nonlinear learning that occurs in the networks of neurons found in nature' (Larose & Larose, 2015, pg. 339).

Abdi et al. (1999, pg. 1) state that '[n]eural networks are adaptive statistical models based on an analogy with the structure of the brain. They are adaptive in that they can learn to estimate the parameters of some population using a small number of exemplars (one or few) at a time.'

As stated by Larson (2012, pg. 620), '[t]he Microsoft Neural Network creates a web of nodes that connect inputs derived from attribute values to a final output.' Each node (the equivalent of a human neuron, also called a processing element) contains two functions: the combination function and the activation function (often referred to as the transfer function). The combination function determines a relative strength (weight and/or importance) for inputs coming into the node and passes that information to the activation function, which determines if the node needs to produce an output. A threshold value that triggers the action of a node can also be used as a form of activation function (Turban et al., 2007, pg. 353).

The nodes are connected to form a network structure (see Fig. 6.4.51). A common neural network structure will have three layers: an input layer, an intermediate (called 'hidden') layer and the output layer. The hidden layer is composed of nodes that accept an input from the previous layer and, after applying functions, convert it into output (Turban, 2007, pg. 350). It should be stated, however, that the presence of the hidden layer is not mandatory. Microsoft's linear regression algorithm is a case of an NN algorithm with a single level of relationships, which according to MacLennan (2008, pg. 373), 'does not necessarily make the logistic regression algorithm a weaker predictor than a full network.' Worth noting is the fact that 'the hidden layer is a very important aspect of a neural network. It enables the network to learn non-linear relationships' (MacLennan et al., 2009, pg. 387).

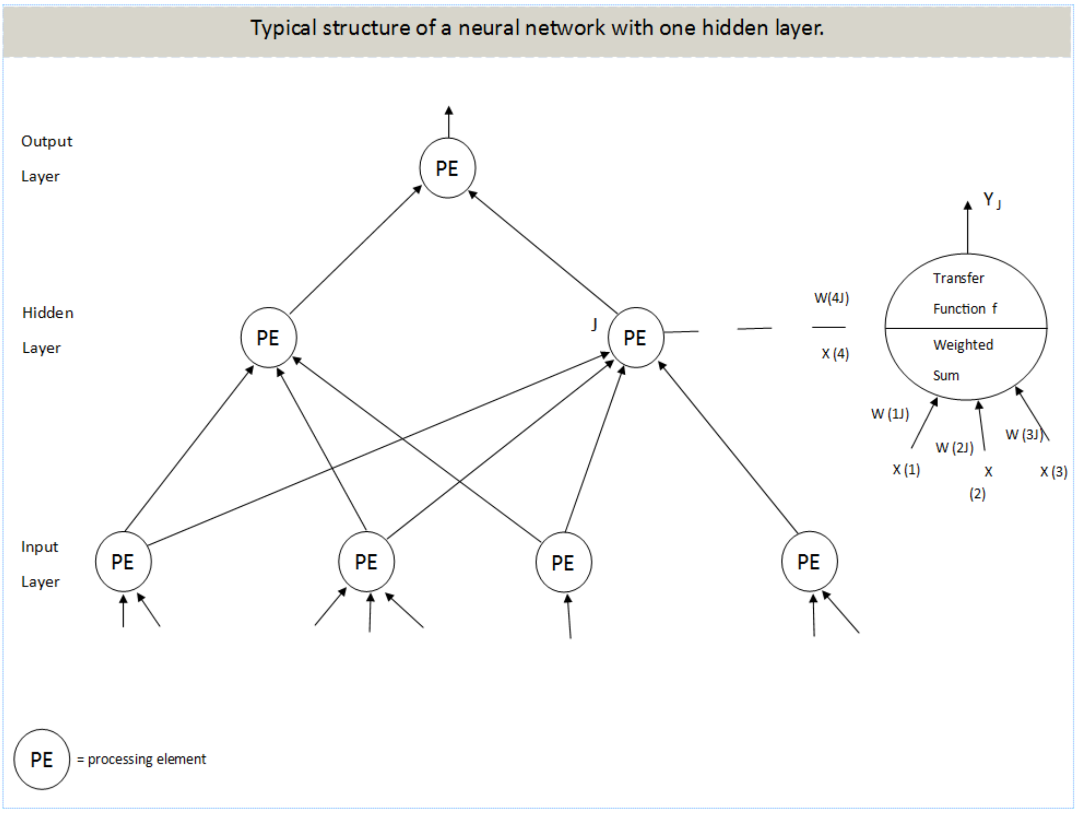


Fig. 6.4.51: Typical structure of a neural network with one hidden layer. [Derived from Turban et al. (2007, pg. 351).]

The NN algorithm, when used for classification purposes, can be either a supervised or non-supervised type of an algorithm, depending on whether the desired outputs (classes) are known and presented to the algorithm. (The algorithm can also be used for the purpose of regression, but, as stated in Table 6.4.1, the regression model is not considered in this research.)

Han et al. (2012, pg. 398) provide some additional characteristics of the NN algorithm that could be used in judging the suitability of the algorithm for solving specific problems. Some of the factors mentioned include the following disadvantages (RQ #5):

- Neural network algorithms takes a long time to train, which can be an issue when using large data sets and/or large number of variables and there is only a limited time window for model building. (After all, the network must consider all possible relationships between inputs and outputs);
- Neural networks uses parameters that are best determined empirically, and parameters are critical for the proper functioning of the algorithm;
- Neural networks are difficult to understand and interpret. (This factor was responsible for the initial slow adaptation of the NN algorithms); and
- It was reported by Andonie (2010, pg. 280) that the NN algorithm is not well suited for use with small data sets: 'Neural Networks have been applied successfully in many fields. However, satisfactory results can only be found under large sample conditions.'

However, the NN algorithm offers the following advantages:

- A high tolerance for noisy data. (This is of very high value in environments where data preparation is challenging);
- An ability to classify patterns on which the algorithm has not been trained. (This is of great value in situations where previously unseen data becomes a part of the input data set);
- Neural networks can be used when there is little knowledge about the relationships between attributes. (While other algorithms can often be used to learn about such relationships, MacLennan et al. [2009, pg. 371]

state that the NN algorithm does a better job of detecting very complex relationships between inputs and outputs); and

- The algorithm works well with discrete as well as continuous data types.

6.4.1 Neural Network Preliminary

With Microsoft's NN's capability to handle both discrete and continuous data types as both input and predicable attributes, there is no need for any special data preparation steps.

With regard to the NN algorithm, the following are the elements affecting performance (RQ #5):

- Normalization and mapping – as pointed out by Abbott (2014, pg. 253), the NN algorithm cannot have missing data and, typically, categorical data is represented numerically through the use of methods described in Section 5.4 – Data Preparation. MacLennan et al. (2009, pg. 393) also point out that a NN requires the values of input variables to be normalized in the same scale, as the larger values are given more weight. The model NN_Model1 created in this section uses the same scale for input as well as output variables (It uses the 'integer' field, as discussed further in the next section.) Discrete variables can be mapped to equal space points on the scale from 0 to 1;
- The topology of the network – this refers to the primary configuration of the hidden layer, as the inputs and outputs are normally specified by the modeler. However, as stated by Han et al. (2014, pg. 400), '[t]here are no clear rules as to the "best" number of hidden layer units. Network design is a trial-and-error- process and may affect the accuracy of the resulting trained network';
- Nonlinearly separable classes – per comments made by MacLennan (2009, pg. 395), the NN algorithm's superiority comes through with problems that take advantage of the non-linear classification used by the NN algorithm (they may have non-linear, and possibly discontinuous, decision boundaries); and
- Algorithm parameters – Microsoft's on-line documentation, 'Microsoft Neural Network Algorithm Technical Reference,' describe the parameters that are supplied to the NN algorithm.

Finally, according the on-line documentation ‘Microsoft Neural Network Algorithm Technical Reference,’ the complexity of a network can be reduced by invoking a method called ‘feature selection’ that reduces the number of ‘considered’ input attributes that are dependent on the values set in the parameters MAXIMUM_INPUT_ATTRIBUTES and MAXIMUM_OUTPUT_ATTRIBUTES.

6.4.2 Neural Network Model

The purpose of NN_Model1 is to illustrate the use of the NN algorithm’s classification abilities. However, as stated by MacLennan (2009, pg. 382), the NN model is slightly different from the remaining models considered: ‘The Neutral Network viewer is different from other Microsoft data mining content viewers in the sense that it is mainly prediction-based. It does not display the information derived from the model content schema row sets, and there is no graphical display of the trained neural network’s layout.’

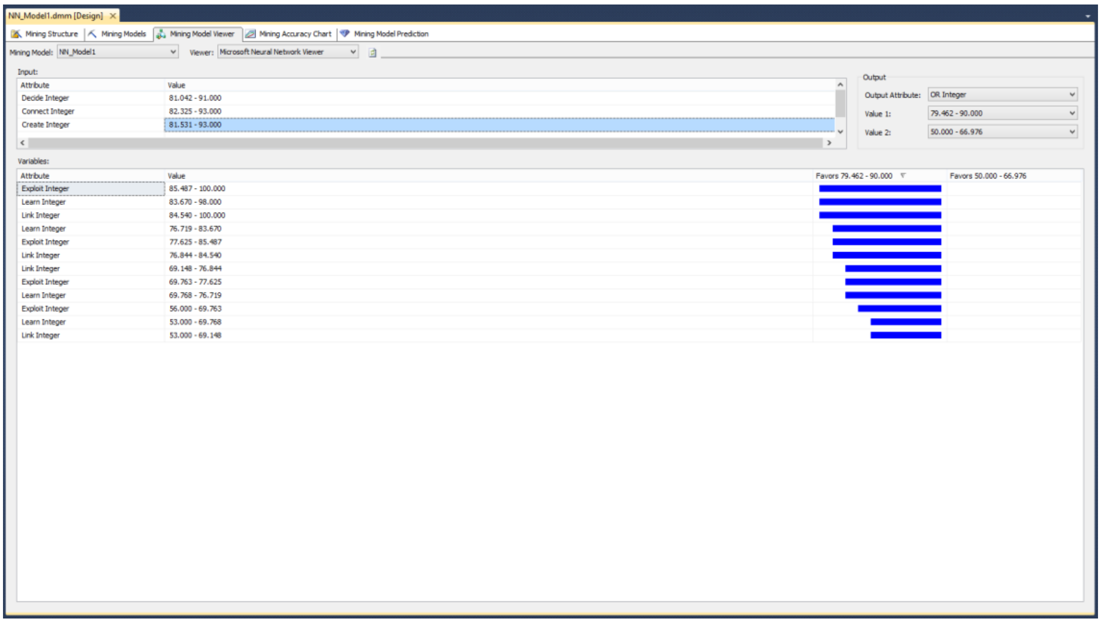
The model used in this research was built using the DM Wizard, which used the previously described data source and held 30% of the data aside for testing. The resultant DM model, NN_Model1, based on the neural network algorithm, used the structure presented in Appendix XI and in figures A8.29 – A8.31 in Appendix VIII. The algorithm’s parameters are presented in Appendix XII.

As seen in Fig. 6.4.52, the display associated with the NN model is very limited when compared to the display available for other mining algorithms, and the appearance, as well as performance, of the NN algorithm is controlled by the parameters (MacLennan, 2009, pg. 396) (RQ #5). Note that, because of the specific need of the NN algorithm for a large data set (Andonie, 2010, pg. 280), and the focus of this research on the illustration of the use of the DM algorithms, changing the default values provided no benefit, even for illustrational purposes (RQ #5).

Fig. 6.4.52 shows a single-tab viewer, the purpose of which is to display the impact of the input attribute and value on the predictable (output) variable; it has three parts. (The viewer is similar to previously presented NB’s attribute discrimination tab.)

The top-left part constitutes an input area that accepts the values of the input attributes that are related to the predictable states. The top right area is the area for output selection: any two values of an output variable are selectable from the drop-down list. Finally, the center pane displays the impact of the attribute and its value on a predictable (output) variable.

It can be seen in Fig. 6.4.52 that, for the three selected input variables ('Decide Integer', 'Connect Integer' and 'Create Integer'), along with the selected highest range of values associated with the three input variables, and the selected output variable's ('OR Integer') values, the three top-most entries ('Exploit Integer, 'Learn Integer' and 'Link Integer', the remaining three competence areas) have a score of 100 in favor of the 'maximum value' (between 79.462 and 90.000) for the output variable (confirming that, if an organization achieves the highest scores in all six competence areas, then its 'OR Integer Score' will also be the highest [RQ #3].)



Mining Model: NN_Model1

Viewer: Microsoft Neural Network Viewer

Input:

Attribute	Value
Decide Integer	81.042 - 91.000
Connect Integer	82.325 - 93.000
Create Integer	81.531 - 93.000

Output

Output Attribute: OR Integer

Value 1: 79.462 - 90.000

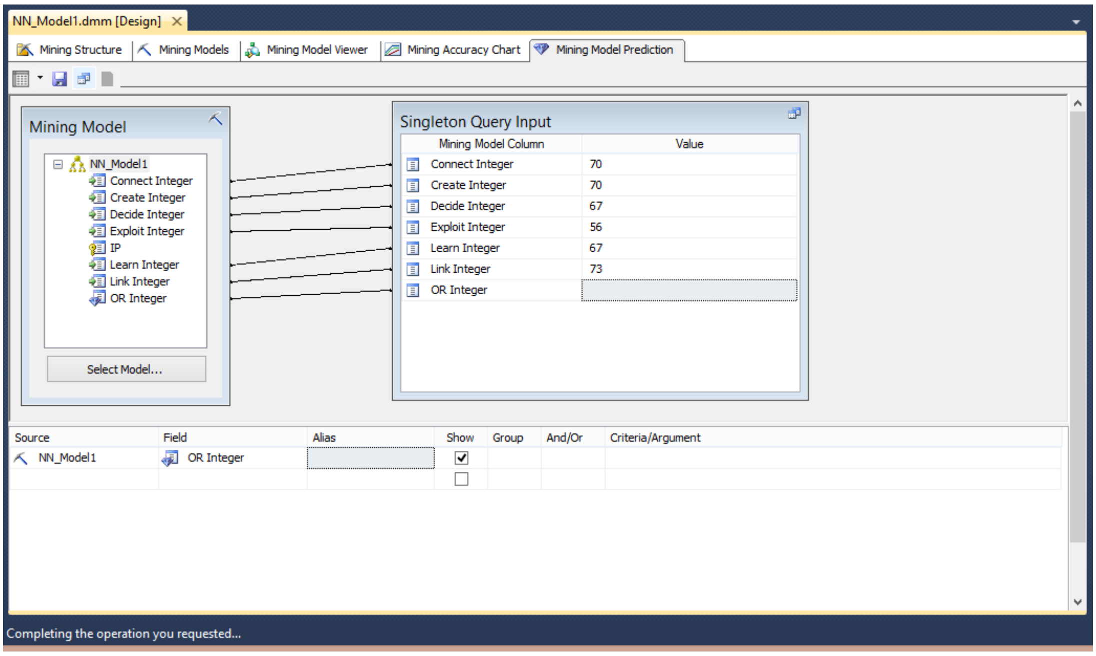
Value 2: 50.000 - 66.976

Variables:

Attribute	Value	Favors 79.462 - 90.000	Favors 50.000 - 66.976
Exploit Integer	85.487 - 100.000		
Learn Integer	83.670 - 98.000		
Link Integer	84.540 - 100.000		
Learn Integer	76.719 - 83.670		
Exploit Integer	77.625 - 85.487		
Link Integer	76.844 - 84.540		
Link Integer	69.148 - 76.844		
Exploit Integer	69.763 - 77.625		
Learn Integer	69.768 - 76.719		
Exploit Integer	56.000 - 69.763		
Learn Integer	53.000 - 69.768		
Link Integer	53.000 - 69.148		

Fig. 6.4.52: The 'Mining Model Viewer' for the neural network

Finally, as mentioned above, with the NN algorithm tailored for the prediction, the final step was to create the predictive model, using the same singleton query and the data taken from the tbl_NBModel2_Predict table as was the case with the NB-based model (but, this time, the data was taken from the 'Integer' columns). Fig. 6.4.53 illustrates the construction of the predictive model:



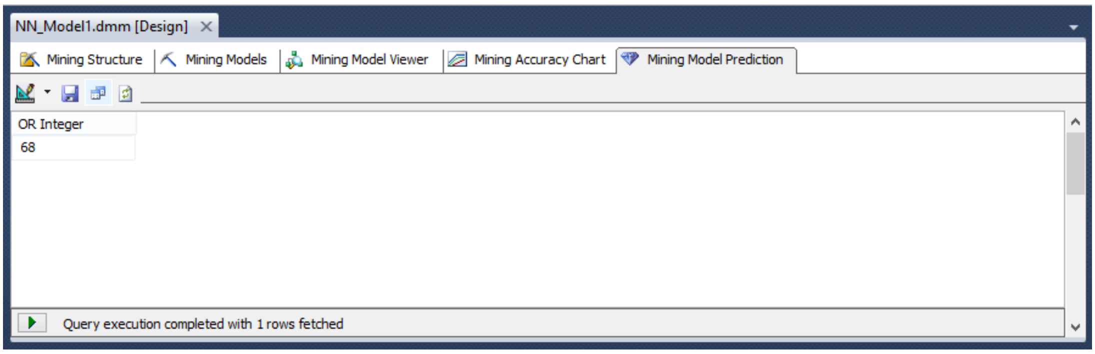
Mining Model Column	Value
Connect Integer	70
Create Integer	70
Decide Integer	67
Exploit Integer	56
Learn Integer	67
Link Integer	73
OR Integer	

Source	Field	Alias	Show	Group	And/Or	Criteria/Argument
NN_Model1	OR Integer		☑			
			□			

Completing the operation you requested...

Fig. 6.4.53: Predictive NN singleton query

The result of the execution of the prediction query is shown in Fig. 6.4.54. The predicted value of the 'OR Integer' given the input parameters is 68, which is significantly lower than predicted by other mining algorithms, but it is important to bear in mind that the small amount of data cannot make any mining prediction valid. (However, the illustration of the comparison of accuracies of predictions among various algorithms is presented in Section 5.6- Model Evaluation.)



OR Integer
68

Query execution completed with 1 rows fetched

Fig. 6.4.54: The result of the NN predictive model using singleton query

Fig. A8.32 in Appendix VIII illustrates the query used to arrive at the result.

NN_Model1

This NN-based classifier DM model used the continuous data of the integer type for the inputs (six competence areas) and as an output (OR) to the model.

Because of the use of NN algorithm and complexities associated with that algorithm, the NN_Model1 DM model constructed differs significantly in terms of functionality (mimicking the human brain) and displayed output information (as no associated with the model network is displayed). While constructing the NN_Model1 DM model, no input parameter needed to be modified; as such, all of the default values were left in place for all NN parameters.

There are two main areas of interest of the functionalities associated with NN_Model1. One area is the classification and the impact of the competence areas on the predictable OR and the second is the predicting capability of the model, referring to how would one actually go about using the predictive capability.

According to the work of Han et al. (2012, pg. 398) and Andonie (2010, pg. 280), one can expect that a NN-based algorithm would perform better in situations where there exist complex and intricate relationships between competence areas and where the relationships between competence area/s and OR are non-linear (RQ #3). However, because of the very limited data, which was identified by Andonie (2010, pg. 280) as a major obstacle for NN-based algorithms, the actual results obtained in this research (and the results were not the objective of this work) cannot be relied upon (RQ #5). Given this, it is unclear which factor was biggest in the resulting NN_Model1, as the output of the model differs from the output of the NB_Model2. While the models differ in their approaches to measurement, it is interesting to see that, while NB_Model2 identified 'Create', 'Link' and 'Learn' Competence (Fig. 6.2.3.16) as the most important competences affecting OR, the NN_Model1 also lists 'Create', 'Decide' and 'Learn' (Fig. A8.42 in Appendix VIII) as important when performing discrimination analysis (RQ #5).

The NN-based model allows discrimination of any two values of the output variable ('OR Integer') and inspection of the corresponding to those values probabilities of input variables (competence areas) favoring specific input variable (RQ #3). As shown in Fig. A8.42 in Appendix VIII, discrimination between the highest and the lowest 'OR Integers' ranges of scores. In the case of Fig. A8.42, by not selecting any individual competence area, it was possible to make a general comparison across all competence areas to determine which competences and which values within those competence areas are favored in the lowest scored 'OR Integer' and the highest cored 'OR Integer' predictable. (In the specific scenario presented in Fig. A8.42, the values in the ranges 81.553-93.000 of the Create competence and the values of 81.042 – 91.000 of the Decide competence were heavily favored by the highest scoring 'OR Integer' while the values of 55.000 – 68.469 of the Create competence were favored by the lowest scoring 'OR Integer' variable.)

Another, more sophisticated, use for the NN-based DM model is that presented in Fig. 6.4.52, where, in addition to the scenario described above, three specific competences areas (Decide, Connect and Create), along with their highest ranges of values, are utilized. Such a configuration allows for discrimination of the 'OR Integer' output variable for the selected two discriminant values, while, at the same time, providing the probability of favoring remaining input variables. That is, in the evaluation illustrated in Fig. 6.4.52, 'Decide' (with the top most range of 81.042-91.000), 'Connect' (82.325-93.000) and 'Create' (81.531-93.000) competences were specifically selected as the input parameters to consider while keeping the 'OR integer' output attribute set to the lowest (50.000-66.976) and the highest (79.462-90.000) values (RQ #3). Then, the impact of selection on the remaining, unselected input variables ('Exploit', 'Link' and Learn'), in terms of probability of favoring certain range of values, was determined (RQ #3). Such a configuration, while very powerful, illustrates the difficulties in keeping all of the relationships 'straight' and the difficulties that may be encountered should such analysis be attempted using classical statistics.

The predictive component of the NN_Model1 worked similarly to the previously presented NB_Model2, using the same input data but a different input variable data type (instead of the categorical values used in NB_Model2, the integer values were supplied as an input in NN_Model1– Sections 5.4 and 6.4.1 describe the needs and methods of data type conversions for the needs of specific algorithms). The prediction query and the result of the query were presented in Fig. 6.4.52. The NN_Model1 arrived at an 'OR Integer' score of 68. While it is difficult to compare the quality of the results from multiple models when there is an insufficient amount of input data to construct fully functional models, the mining accuracy chart and model tests are typically used to select the best for a given task DM model. A discussion assessing the model's prediction ability is conducted in Section 6.6.

The construction and interpretation of the NN_Model1 was found not to be overly complex. However, treating the inner workings of the algorithm as a 'black box' makes the model testing difficult, as the correctness of the model cannot be easily confirmed (for example, with the use of the sample example from the model and classical statistics) (RQ #5).

6.5 Data Mining: Decision Trees (DT)

The decision trees (DT) algorithm is regarded, perhaps due to its similarities to the "if-then-else' constructs used in business, as one of the easiest algorithms to understand. Larson (2012, pg. 611) states that '[t]he Microsoft Decision Trees algorithm is one of the easiest algorithms to understand because it creates a tree structure during a training process.' The tree structure constructed is, however, influenced by the algorithm's parameters, which may not be so easy to understand. The mining structure is then used to provide 'classification-based' predictions and analysis. (Microsoft refers to the decision tree algorithm as decision trees due to the different 'tree shapes' obtained from a single algorithm, based on the various setting of parameter values [MacLennan, 2019, pg. 236]).

Provost and Fawcett (2013, pg. 63) give the following description of the construction of a generic decision tree algorithm: The data is segmented into a tree-like shape, positioned upside down, with the root at the top. The tree is made up of nodes (there are two types: internal, those having nodes beneath them, and terminal, the leaves). Branches connect the leaves (a binary tree will have at most two branches out of one internal node). The tree truly creates a segmentation of the data, as every data point corresponding to only one path and one leaf. (As stated by Provost and Fawcett, 'each leaf corresponds to a segment, and the attributes and the values along the path give the characteristics of the segment.')

The DT algorithm represents a supervised type of algorithm learning because each leaf contains a target value – that is, the class label is provided at each leaf.

Han et al. (2012, pg. 330) describe decision tree induction, or learning, from the decision tree's class labels. Each internal node denotes a test on an attribute, and a branch represents an outcome of a test on an attribute. To use a decision tree as a classifier, the value of an attribute is tested against the decision tree, and a path from the root to the leaf node is created (the leaf holds the class label). Provost and Fawcett (2013, pg. 67) state that, '[i]n summary, the procedure of classification tree induction is a recursive process of divide and conquer, where the goal at each step is to select an attribute to partition the current group into subgroups that are as pure as possible with respect to the target variable.' Two of the leading algorithms for the construction of decision trees are the classification and regression trees (CART) algorithm and the C4.5 algorithm (Larose & Larose, 2015, pg. 319).

One of the key aspects of the functioning of a DT algorithm is the choice of attribute selection for each node. While the mathematical background that makes up the attribute selection methods is outside of the focus of this work, it is worthwhile to name, for the sake of completeness, some of the key methods used. These include the following (Han et al., 2012, pg. 336):

- Shannon's information gain (purity measure) entropy and variance reduction – selecting the nodes with the highest 'information content';
- Gain ratio – a method that removes the bias of the information gain measure towards selecting attributes that have a large number of values; and
- Gini index – a method that measures the impurity of the training data set using mathematical formula.

6.5.1 Decision Trees Preliminaries

As with other DM algorithms used in this work, the DT algorithm has some practical considerations that must be taken into account.

Abbott (2014, pg. 229) identifies a number of issues to consider (RQ #5):

- Re-trying the model construction by removal of the variable placed originally as the root of a tree. Forcing the reconstruction of the tree as a decision tree can often be suboptimal, as the algorithm only has a single chance to select a node (and it never 'goes back' to consider other options);
- Because it only uses one variable at an internal node (split), the DT algorithm needs a good start as otherwise a less than optimal tree will be built. To assist with building an optimal tree, the modeler should include multivariate features, if known;
- Trees are considered unstable models as, often; even small changes can change how the tree looks or behaves. For that reason, it is important to inspect the 'runner-up' and the winning split in order to understand how valuable the winning splits are;
- Trees are biased toward selecting categorical variables with large number of levels (high cardinality data). If such variables are found, methods described in Section 5.4 need to be applied to the data (mainly binning methods); and
- Single trees, on average, are not as accurate as other predictive algorithms, primarily because of 'greedy' forward (one chance) variable selection.

Additional requirements for using a DT algorithm have been identified by Larose and Larose (2015, pg. 319) and include the following (RQ #5):

- A DT algorithm, representing supervised learning, needs to have pre-classified target variables. (In case of this work, this will indicate a Boolean 'yes/no' type of variable that indicates if an organization responding to the questionnaire is resilient or not, based on some value set for the variable indicating resilience);
- The training data should contain all possible choices and be varied to ensure the algorithm 'sees' as many possible combinations of answer and results as possible; and
- The target variable must be of the discrete type, so that it can be clearly classified if the obtained values does or does not belong to a given class.

As is the case for all models presented in this research, DT model construction takes place within the Microsoft environment.

6.5.2 Decision Trees Model

DT_Model1 was created in order to illustrate the use of the classification capabilities of the DT algorithm. For the purpose of the creation of this model, without affecting all previously created models, a copy of the table used in all previous model construction was made, saved under the new table name of tbl_DM_KM_OR_RGU_DecTree. Then, a new column, IsOR of binary type, was added to the new table for the purpose of using this binary field (resembling many predictions made in real life that are of a yes/no type of answer).

The value in the newly added field IsOR was set according to the following formula: If the value in the field ORInteger was 80 or higher, then the value in the filed IsOR was set to 1 (to indicate a reply from the "resilient organization"). For all other values in the ORInteger field, the value in the IsOR was set to 0 (to indicate the "non-resilient" organization's reply). (Clearly, the values for the formula have been chosen somewhat arbitrarily, and a different 'cut-off' value could have been used. However, to the author of this thesis, an achievement of 80% or better in terms of OR constitutes the passing grade; thus the selection of cut-off value.)

Because of the intention to use a new table, a new data source view (called RGU_DInfSC01) had to be established. Later, this data source view was used in the construction of the DT_Model1 model, and the data view table (tbl_DM_KM_OR_RGU_DecTree) became the source for the analysis.

The DT algorithm parameters, per MacLennan (2009, pg. 256), are presented in Appendix XII.

Several unsuccessful models were attempted before the creation of the successful model. The factors in successful model creation appear to be the settings of the parameters MIN_SUPPORT, SPLIT_METHOD and SCORE_METHOD. Per earlier notes in the prior section, the data type of the output variables, as there are two in this model, needed to be of the discrete type.

Because of the very small number of input records, the default value of 10 for MIN_SUPPORT produced no tree (and no dependency network) (RQ #5). It was determined that, in order for the algorithm to produce meaningful results for the purposes of the illustration model, the value of the MIN_SUPPORT parameter had to be set to no more than 2. With regard to the other two parameters responsible for the performance, the value of SCORE_METHOD parameter and SPLIT_METHOD parameter had to be set to 1 in both cases, as, otherwise, the following issues occurred (RQ #5):

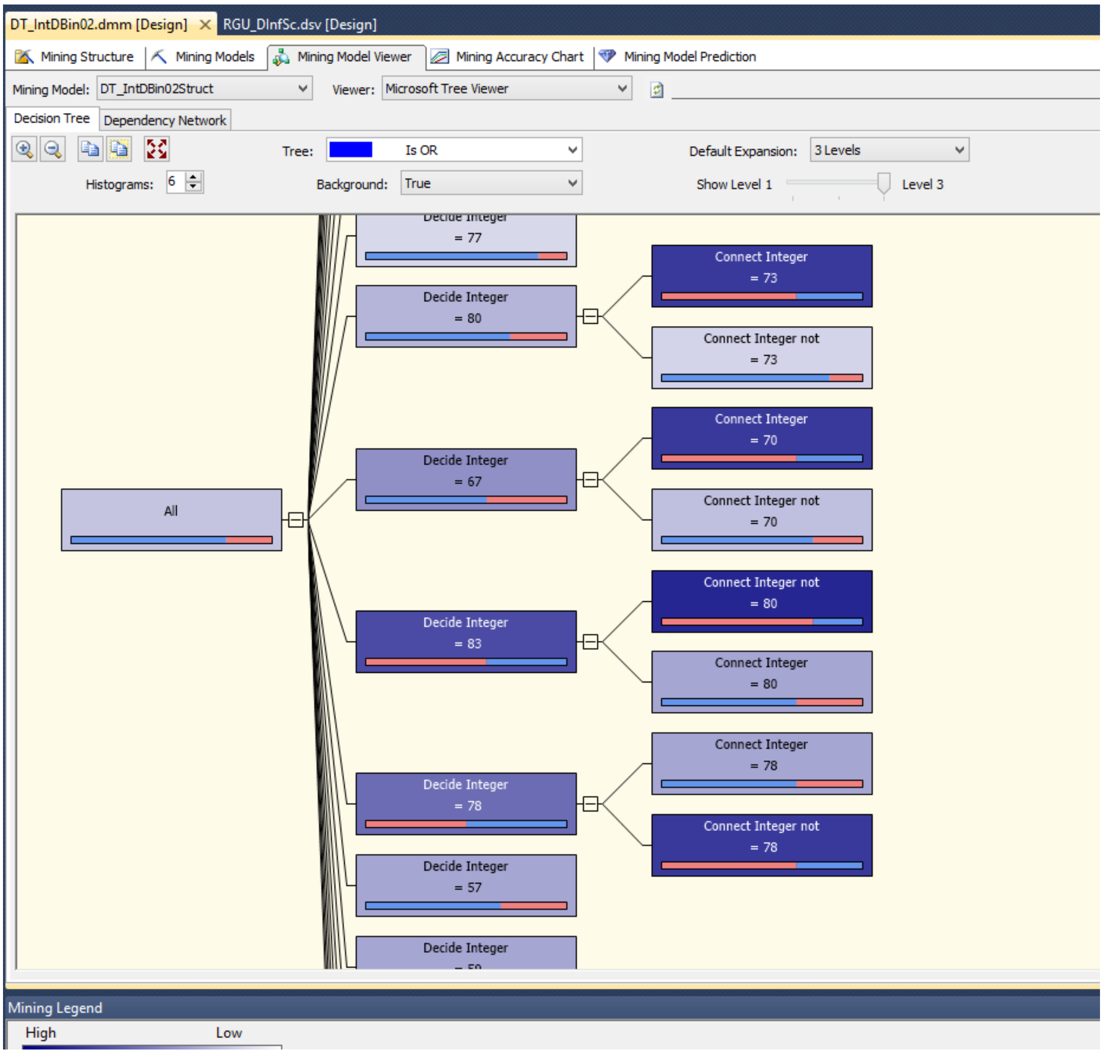
- SCORE_METHOD = 1 and SPLIT_METHOD = 2 – the tree created was not of a binary form and considered each output value individually in constructing the tree. (The outcome of using these parameter values is shown in Fig. 6.5.55, and the outcome is the 'unmanageable' number of splits.)
- SCORE_METHOD = 1 and SPLIT_METHOD = 3 – similar outcome to that described above for the value of SPLIT_METHOD = 2.
- SCORE_METHOD = 2 and SPLIT_METHOD = 2 – obtained an error that the model could not be built using these parameters.
- SCORE_METHOD = 2 and SPLIT_METHOD = 3 – obtained an error that the model could not be built using these parameters.
- SCORE_METHOD = 3 and SPLIT METHOD = 2 – the constructed tree found no splits. (The constructed tree consisted of a single 'All' node.)
- SCORE_METHOD = 3 and SPLIT METHOD = 3 – the constructed tree found no splits. (The constructed tree consisted of a single 'All' node.)

The structure of the DT_Model1 is represented in Appendix VIII, Fig. A8.37 and Appendix XI.

The decision trees model, DT_Model1, was constructed using the data view 'RGU_DInfSC01' created for this model. As before, the model was constructed using the DM Wizard. The default 30% of the data was set aside for model testing. All of the DT_Model1 construction steps described above are illustrated in Appendix VIII, Fig. A8.33 – A8.39.

A new approach with respect to the DT_Model1 is the use of 2 predictable variables: 'Is OR' and 'OR Int Discretized'. The 'Is OR', the 'yes/no' type of the variable, was added to illustrate a typical use of the DT algorithm in the field:

determining if the predicted variable is or is not of a specific type (similar to a loan officer asking the question ‘is the applicant credit-worthy?’). The second variable, ‘OR Int Discretized,’ was used to be consistent with the previous approaches used in this work; instead of arriving at a yes/no answer, it allows arriving at a numeric score: ‘OR Score’ (RQ #3).



Mining Model: DT_IntDBin02Struct

Viewer: Microsoft Tree Viewer

Tree: Is OR

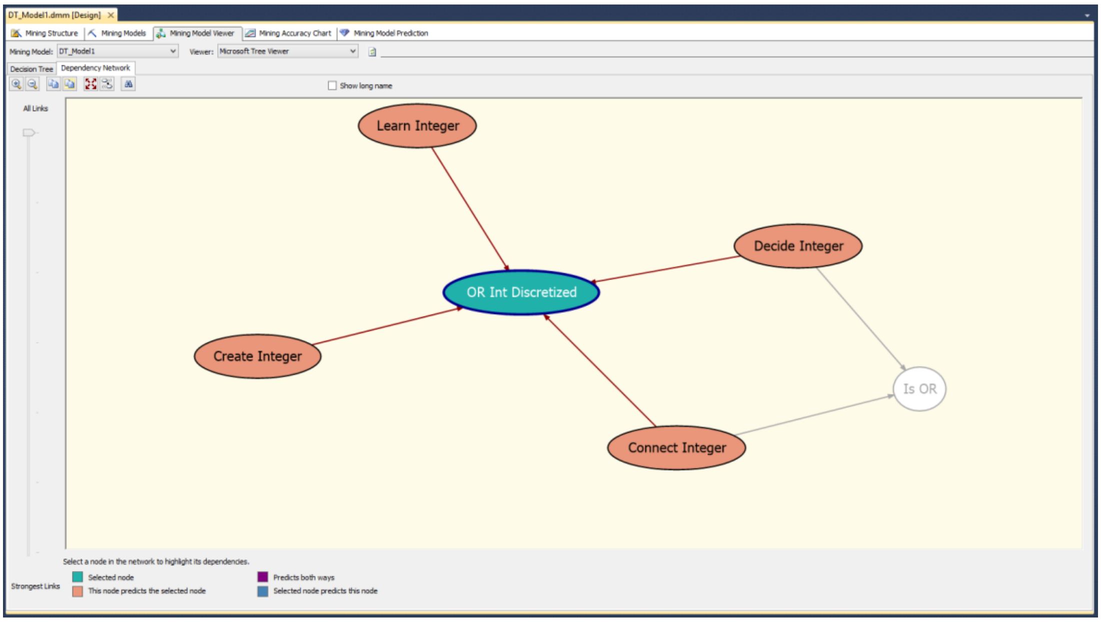
Default Expansion: 3 Levels

Histograms: 6

Background: True

Fig. 6.5.55: DT algorithm outcome with SPLIT_METHOD set to value other than 1

A As previously discussed, the mining model viewer shows the two predictable variables in the dependency network diagram. ‘OR Int Discretized’ is the selected output variable in Fig. 6.5.56. (Interestingly, the links between the two output variables differ [RQ #3]. Clearly, the limited amount of data and the different ‘granularity’ of the data in the two output variables can explain the differences in linkages.) The strongest link in the entire diagram, presented when the slider is moved to the ‘Strongest Links’ position (which is not shown in the diagram in Fig. 6.5.56) is between the ‘Learn Integer’ input variable and the ‘OR Int Discretized’ output variable.



Mining Model: DT_Model1

Viewer: Microsoft Tree Viewer

Select a node in the network to highlight its dependencies.

Fig. 6.5.56: DT dependency network

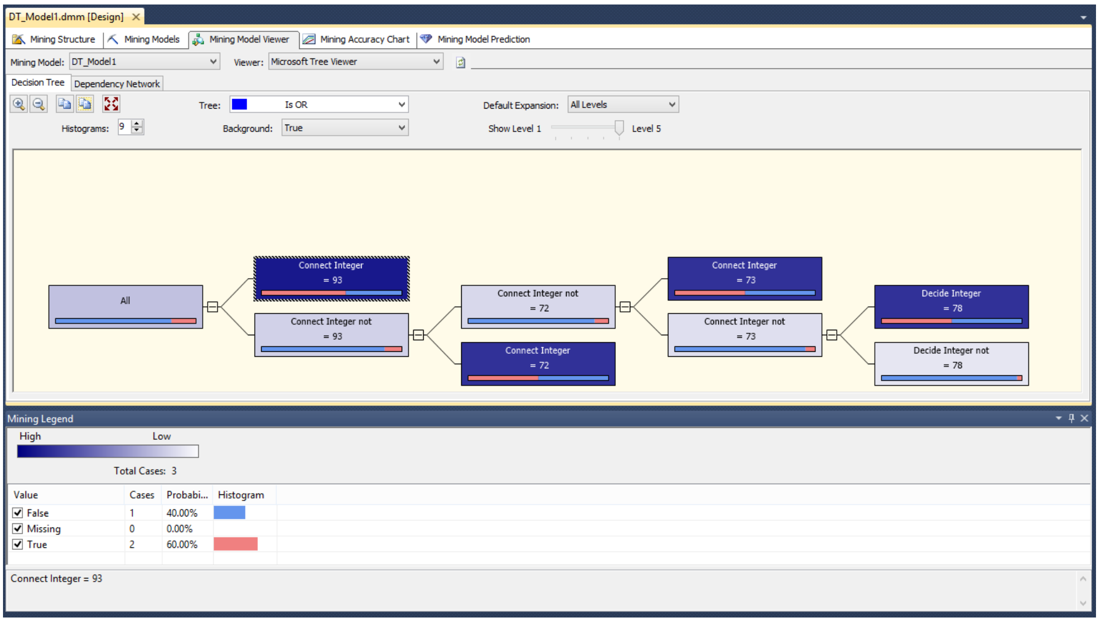
The DT tab of the mining model viewer shows the resulting DT for the ‘Is OR’ predictable variable, shown in Fig. 6.5.57.

In the top section of the display, there are number of parameters that control display. The ‘Tree’ drop-down box, set to ‘Is OR’ in Fig. 6.5.57, is the area where the selection of the output variable is made. Immediately below, there is another drop-down box, labeled ‘Background’. This ‘Background’ selector controls which value of the output variable selected in the ‘Tree’ selector to build a decision tree for. In the case of Fig. 6.5.57, the ‘Background’ value is set to ‘True’, to build a tree for the case where ‘Is OR’ = ‘True’.

Located to the right of the ‘Tree’ and ‘Background’ selectors is the area controlling the height of the displayed tree. In Fig. 6.5.57, all tree levels are shown.

At the bottom of Fig. 6.5.57, the mining legend can be seen, with the darker color indicative of the presence of the specific value (selected as ‘Background’) in a given node. The small horizontal bar within each node represents the ‘ratio’ of the number of entries that have the desired value of output variable to the number of entries that have different values. The value in the box labeled 'Histograms' controls the number of (states) colors to display in the small horizontal bar present within each node. (For the 'Is OR' variable, the value for the 'Histogram' can be set to two, to represent 'True' and 'False,' as there are no other values or missing entries.)

At the bottom of the screen, the mining legend for the selected node is displayed. Inspection of the mining legend for the leaf node labeled 'Connect Integer = 93' shows that the node contains a total of three cases. Two, or 60% of the population in that node, has a value of 'True' and one, or 40%, has a value of 'False.' The 60% – 40% ratio is displayed as a small horizontal bar, with the 'True' cases having 'pink' color and the 'False' cases the blue color.



Mining Model: DT_Model1

Viewer: Microsoft Tree Viewer

Tree: Is OR

Default Expansion: All Levels

Histograms: 9

Background: True

Show Level 1 Level 5

Total Cases: 3

Value	Cases	Probabi...	Histogram
☑ False	1	40.00%	🟦
☑ Missing	0	0.00%	
☑ True	2	60.00%	🟥

Connect Integer = 93

Fig. 6.5.57: Classification tree for 'Is OR' variable and value True

Interpretation of the classification presented in Fig. 6.5.57 can be presented as follows (RQ #3):

1. The algorithm begins tree construction with a node, labeled 'All,' that contains all of the data elements to be used in the tree's construction (it is important to bear in mind that the algorithm, being of the supervised type, knows the value of 'Is OR' for all sets of inputs, with a set being a single value in each input and output variable). The decision node 'All' is of relative low blue intensity, indicating no great concentration of the

values ‘True’ with relation to ‘False’. Clicking on the ‘All’ node and reading off the mining legend’s values, it can be seen that the node has 32 cases. Twenty-seven of these result in value ‘False’ and are marked in blue in a horizontal histogram inside the node; 5 cases resulted in value ‘True’ and are marked with the color pink. Therefore, the histogram inside the ‘All’ node has 82% probability of blue, or ‘False,’ values and 18% probability of pink, or ‘True,’ values. The first split is made on the ‘Connect Integer.’ (‘Connect Integer’ was shown as one of the key influencers for the ‘Is OR’ variable in the dependency network diagram.)

2. The first split on the ‘Connect Integer’ input variable results in one of the nodes being a leaf node (the node labeled ‘Connect Integer = 93’) and one a decision node (with values not equal to 93 in the ‘Connect Integer’ variable). The leaf node, when selected with the mouse, shows the following characteristics: It contains 3 cases for ‘Is OR,’ where one case has a value of ‘False’ and is given a 40% probability of occurring and two cases have values of ‘True’ and 60% probability. Therefore, the first ‘classification path’ for ‘Is OR’ is the value ‘Connect Integer’ equals 93, which gives a chance of 60% for the ‘True’ value of the output variable. The other, the decision node, has 29 cases with 26 being of value ‘False’ (87% probability) and three cases of value ‘True’ (13 % probability).
3. The second split, at the node labeled ‘Connect Integer not = 93’, produces the second leaf node, ‘Connect Integer = 72,’ and another decision node: ‘Connect Integer not = 72’. The second leaf node contains only one case with the value of ‘True’ (meaning that, after this split, there were only two more cases with the value of ‘True’) and another decision node. Both values are given a 50% probability of occurring, which is also indicated by an equal split between the blue and pink on the horizontal histogram. (At this stage, the classification rules are as follows: if an input variable has ‘Connect Integer’ equal to 93, then there is a 60% chance that the ‘Is OR’ for that variable will have the value ‘True’. On the other hand, if the ‘Connect Integer’ does not have a value of 93 (and there is an 87% probability of that happening), then, if the ‘Connect Integer’ has value of 72, there is a 50% probability that those variables with a value of 72 will have ‘Is OR’ equal to ‘True’.

The process continues, and as can be seen in Fig. 6.5.57, the remaining two cases with the 'Is OR' value of 'True' can be found in the leaves 'Connect Integer' = 73 and 'Decide Integer' = 78.

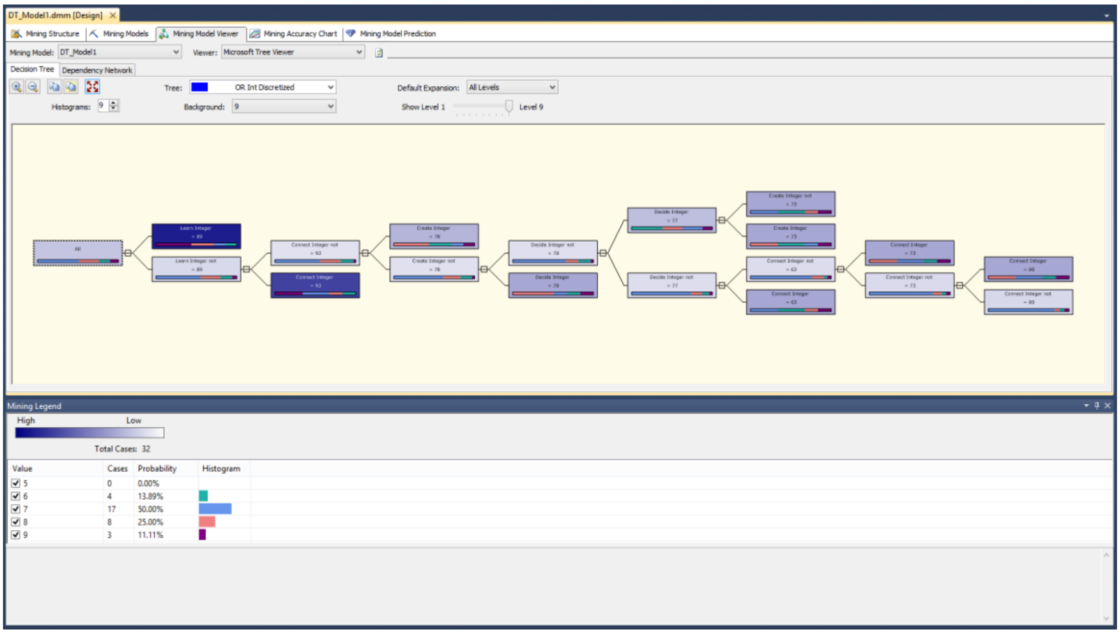
The next figure, 6.5.58, shows the tree constructed for the other predictable variable 'OR Int Discretized,' attaining the value of 9 – the display was generated after using the 'Size To Fit' option for the display, due to the large size of the resultant tree.

Because the output of the 'OR Int Discretized' variable, in the scenario used in this research, can take a value between 5 and 9, the histogram now contains six colors instead of two, as was the case with the 'Is OR' Boolean variable.

The mining legend for the 'All' node shows, as before, the total of 32 cases; however, for the 'OR Int Digitized' variable, there was a total of 3 cases with the value of 9.

As can be seen in Fig. 6.5.58, all 3 cases were present in the first two leaf nodes. The 'Learn Integer' = 89 node contained 2 cases, hence the node having the darkest color (with an estimated probability of 11% for a '9' value to appear). The slightly lighter-colored leaf node 'Connect Integer' = 93 contained the remaining single case when 'OR Integer' = 9 with the probability of occurring equal to 33%. In the illustration classification, it can be seen that, if the input variable does not have the 'Learn Integer' equal to 89 or it does not have the 'Connect Integer' equal to 93, then one cannot expect to find the value of '9' in any other 'situation'. (So, in the notation used by the DT, the rules for arriving at the value of '9' in the output variable are as follows: [Learn Integer = 89] OR [Learn Integer not = 89 AND Connect Integer = 93].)

The DT prediction model is presented in Fig. 6.5.59, and the query itself is presented in Appendix VIII, Fig. A8.40. In the DT-based prediction, the same data as for other predictive models was used. In this model, however, the goal is to obtain a prediction regarding two output variables at once: the 'Is OR' and the 'OR Int Discretized.' (That is, the goal is to discover if the organization responding to the questionnaire's question is resilient and, regardless of the answer to the first question, to discover its resilience score [RQ #3]) For the purpose of prediction, the singleton query (single values selected at the input screen) was used.



Mining Model: DT_Model1

Viewer: Microsoft Tree Viewer

Tree: OR Int Discretized

Default Expansion: All Levels

Histograms: 9

Background: 9

Show Level 1 Level 9

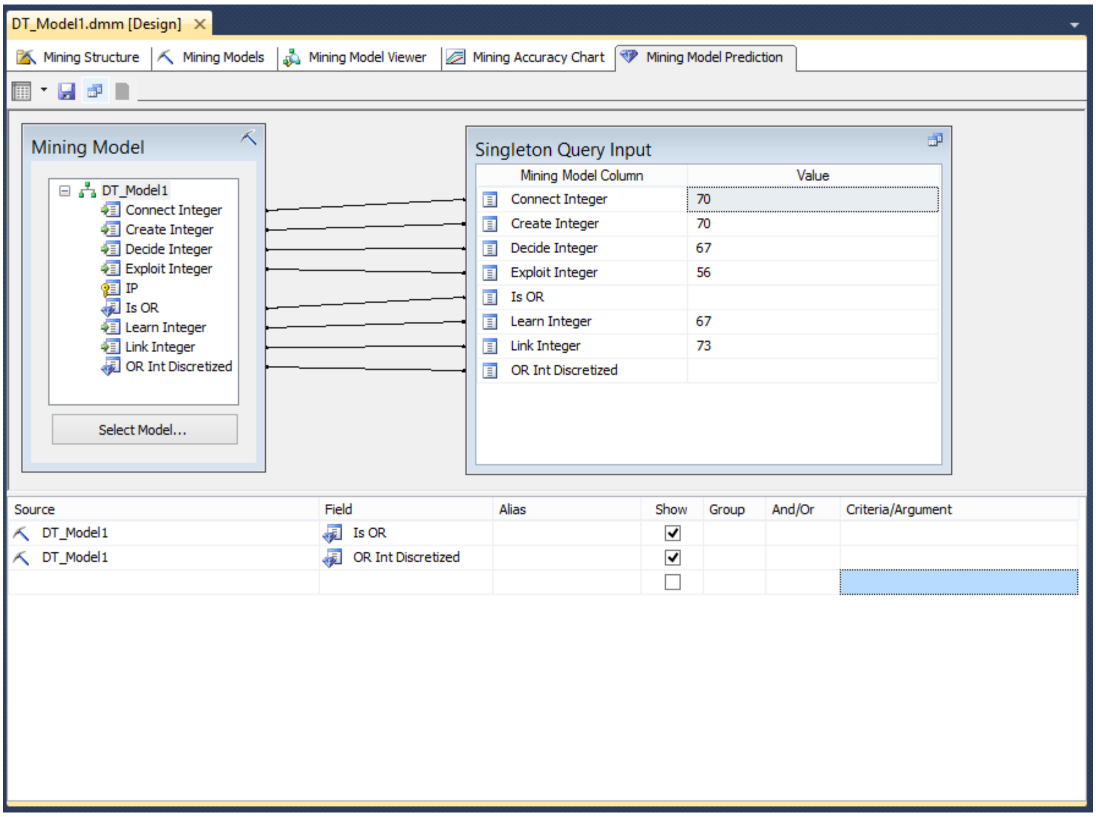
Total Cases: 32

Value	Cases	Probability	Histogram
☑ 5	0	0.00%	
☑ 6	4	13.89%	🟩
☑ 7	17	50.00%	🟦
☑ 8	8	25.00%	🟥
☑ 9	3	11.11%	🟪

Fig. 6.5.58: DT for ‘OR Int Discretized’ variable attaining value 9 with the ‘Learn Integer’ = 89 node selected

The outcome of the prediction is shown in Fig. 6.5.60. As can be seen, the results indicate that the firm responding to the questionnaire is not resilient, based on the somewhat arbitrarily set resilience level (‘Is OR’ has value of ‘False’), and its resilience score is 7 (out of 10).

Clearly, very limited data does not allow drawing any conclusions from this predictive exercise, but the issue of prediction accuracy, within the context of all of the algorithms used, is presented in the next section (RQ #5).

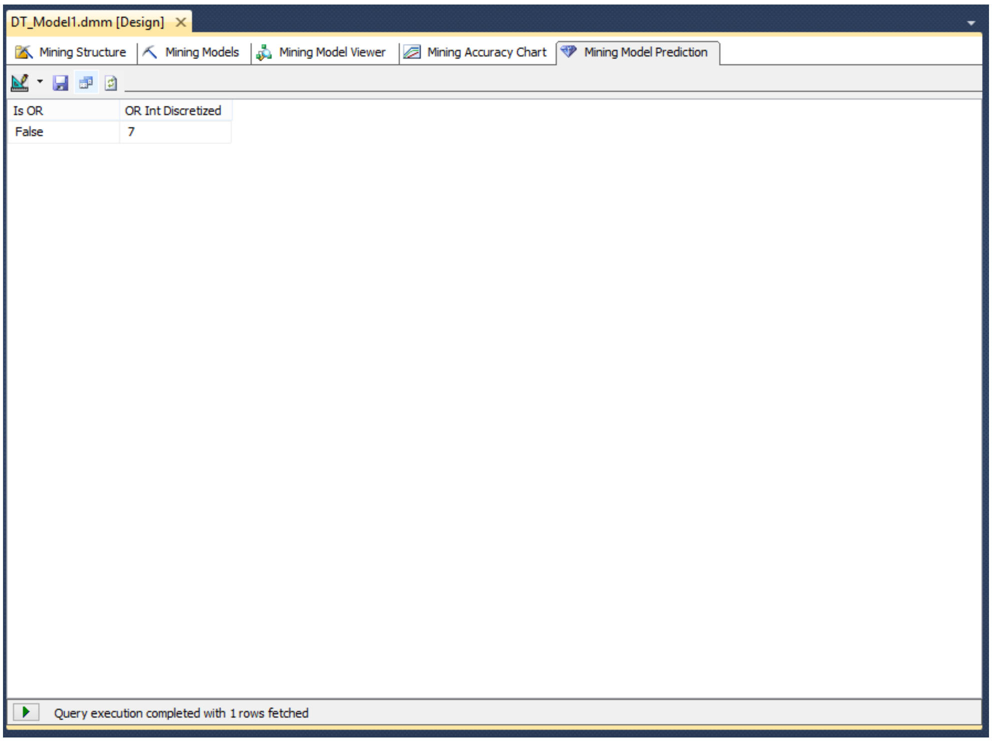


DT_Model1
- Connect Integer
- Create Integer
- Decide Integer
- Exploit Integer
- IP
- Is OR
- Learn Integer
- Link Integer
- OR Int Discretized

Mining Model Column	Value
Connect Integer	70
Create Integer	70
Decide Integer	67
Exploit Integer	56
Is OR	
Learn Integer	67
Link Integer	73
OR Int Discretized	

Source	Field	Alias	Show	Group	And/Or	Criteria/Argument
DT_Model1	Is OR		☑			
DT_Model1	OR Int Discretized		☑			
			□			

Fig. 6.5.59: DT prediction model.



Is OR	OR Int Discretized
False	7

Query execution completed with 1 rows fetched

Fig. 6.5.60: DT prediction results

The decision trees-based model uses integer types of input variables (competence areas) and two types of output variables: the discrete binary type (holding values 1=Yes/0=No and called 'IsOR') and the discrete integer ('ORIntDiscretized'). The output variables are used one at a time, with the 'IsOR' variable being used to determine if the prediction result returns 1, meaning that the organization under consideration is resilient or returns 0, for a non-resilient organization (RQ #3). The source table tbl_DM_KM_OR_RGU was replicated for the purpose of the decision trees algorithm (to ensure that models created previously were not affected by addition of the 'IsOR' column).

With the DT model, the 'IsOR' column become the supervisor variable for a given questionnaire reply. The value of the field was set to 1 (or Yes, meaning that the response represents a response from a resilient organization) if the value of the 'ORInteger' was >= 80. Otherwise, the value of the field was set to 0 (not resilient). The choice of value for the supervisor variable was somewhat arbitrary, but, in real life, scenario analysis would need to be conducted in order to determine the appropriate ORInteger value that would represent a resilient organization.

Perhaps the most challenging aspect of the entire process of constructing the DT_Model1 model was the proper setting of the algorithms parameters (RQ #5). The parameter setting and the challenges encountered related to those settings, including unsuccessful model construction, were described in Section 6.5.2.

One of the first results obtained from the model construction was the network diagram (shown in Fig. 6.5.56). The diagram allows viewing the dependencies between the competence areas and the OR (RQ #3). In the case of the diagram constructed for DT_Model1, 'Learn', 'Create', 'Connect' and 'Decide' competences predicted 'ORIntDiscretized,' while 'Connect' and 'Decide' predicted the second output variable, 'IsOR'. (Note that there appears to be somewhat of an agreement between the models in terms of the key competences, as NB_Model2 identified 'Create', 'Link' and 'Learn' competences (Fig. 6.2.3.16) as affecting OR the most, while the NN_Model1 lists 'Create', 'Decide' and 'Learn' (Fig. A8.42 in Appendix VIII), as the most important competences when performing discrimination analysis. Clearly, the limited amount of input data does not allow for drawing any conclusions. (RQ #5)

In addition to the network diagram, two decision trees where produced, one per output variable.

When the DT_Model1 utilized the ‘IsOR’ field for prediction, the outcome of the prediction was an answer to the question of whether or not the organization submitting the responses was resilient or not (RQ #3). In order to determine the ‘OR Score,’ as was the case with the models already discussed in this section, field ‘ORIntDiscretized’ had to be used.

The decision tree produced for the ‘IsOR’ output variable was constructed for one of the two possible outputs of the’IsOR’: true or false. (The trees constructed are shown in Appendix VIII, Fig. A8.43 and Fig. A8.44.). The tree constructed in this fashion allows for determination of the importance of the competence areas, based on the internal level node splits, with the most influential competence (‘Connect’) being closer to the ‘All’ node. The information presented within each node (node’s background color intensity = % of the selected, true or false, values of ‘IsOR’ variable, while histogram = break down of the presence of each value, ‘true’ or ‘false’ of the ‘IsOR’ output variable) allows for easy interpretation of ‘OR’ for the constructed tree.

Presented to investigate the ‘ORIntDiscretized’ output variable display (Fig. A8.45 and A8.46 in Appendix VIII) while it provides similar benefits as in case of ‘IsOR’ variable in terms of classifying firms as resilient or not, the internal nodes evaluate competence area with respect to the content of specified input value (5 to 9) instead true or false. Using the discretized output value for OR (and the appropriate parameter values described in Section 6.5.2), the constructed decision tree has more levels than the tree for ‘IsOR’, something that was expected due to the larger input set (five as opposed to nine). What is interesting, however, is that both trees use different internal nodes. The ‘IsOR’-based tree mainly used the ‘connect’ competence with the ‘decide’ competence at the leaf level, while ‘ORIntDiscretized’ initially used the ‘Learn’ competence, followed by ‘connect’, ‘create’, ‘decide’, and ‘connect’ and ‘create’ as the final leaves (RQ #5). The difference in construction can perhaps be attributed to the small data set, but this example indicates the need, discussed by Abbott (2014, pg. 229), to re-try model construction by substituting different variables as the tree root until good stability has been achieved (RQ #5).

While the DT-based models are relatively easy to understand, they are not overly popular in the field for a number of reasons; the main reasons are the complexities of the parameter setting as well the fact that a single attempt to find the best variable for internal node often leads to less than optimal trees (Abbott, 2014, pg. 229). This research found that setting parameters is complicated and requires some trial and error and that single attempt of split at the internal node selection applied, as expected, but no comment can be made about the optimality of the tree due to the limited amount of data (RQ #5).

6.6 DM Model Evaluation

This section continues the discussion started in Section 5.6 which introduced elements of the model evaluation common to all the DM models.

The DM Wizard, during the process of creating a DM model, splits the input data set into two sets: the training data set and testing data set (Appendix VIII, Fig. A8.21). The training dataset is used to build the DM model, and the testing dataset is used to check the model's accuracy. (In all of the models used in this research, the default, 30% of the data, was allocated for testing.)

As stated by Janus and Misner (2011, pg. 357), the (mining) accuracy chart, a feature of the SLQ Server 2012 platform used in this research, can be used to evaluate a predictive model, provided the model is not based on a time series or association rules algorithm, as the chart shows the improvements in accuracy of the prediction as the population size increases. The following figures (6.6.1 – 6.6.4) indicate the outcomes of the creation of the mining accuracy, or lift, chart for each of the DM algorithms used. The type of the chart presented depends on the data type of the target variable: a different chart is presented for continuous and different for discrete target variables.

Fig. 6.6.2 represents a lift chart for the Naïve Bayes model discussed in Section 6.2. 'OR Int Discretized' was chosen as the selected predictable column to ensure compatibility with that used in other models, and '9' was selected as the value to predict. (This selection is illustrated in Appendix VIII, Fig. A8.41.)

The charts presented in Figures 6.6.1 and 6.6.2 are examples of a typical lift chart presented for discrete type of output variable. (In the case of the

DT_Model1, the lift chart for which is displayed in Fig. 6.6.1, the output variable is 'OR Int Discretized,' which was specified during the model construction to be of the 'Discrete' type. In the case of DT_Model1, the output variable was also 'OR Int Discretized', specified to be of 'Discrete' type.) In both cases, the value of '9' was chosen as the value to predict.

The outcome of the lift charts, shown in figures 6.6.1 and 6.6.2 suffers from the extremely small amount of input and testing data and does not allow the performance of the algorithms to be properly evaluated.

The problems caused by the lack of data are to be expected in lift charts, as stated by Larose and Larose (2015, pg. 463): 'Lift is a function of sample size….' The authors define the lift as 'the proposition of true positives, divided by the proposition of positive hits in the data set overall:

Lift = Proposition of true positives / proposition of positive hits, which is equivalent to:

$$\text{Lift} = \frac{\dfrac{\text{True positives}}{(\text{False positives} + \text{True positives})}}{\dfrac{(\text{False negatives} + \text{True positives})}{\text{Sample size}}}$$

Therefore, as a result of the very small sample size used in this research (32 elements, not counting the elements 'excluded' for testing) and the six bins (corresponding to the six competence areas), there are not enough elements for the ratio to exceed the 'ratio of the random line,' so there is no lift (lift occurs when the performance chart occurs above the random line, creating lift from a random line). Hence the unexpected shape and 'below the random line' location of the performance curve.

When looking at Fig. 6.6.1, a few points stand out as out of the ordinary. First, the performance of the DT_Model1 model (performance curve), represented by a red line, is, for the most part, below the random guess, or the 45 degree blue line. Second, the shape of the performance curve on the lift chart has 'very long periods' of the straight lines with only three 'vertices'; clearly, this is an indication of the presence of something unexpected with relation to the testing data, namely the unusually small amount of testing data. (Typically, the 'performance curve' of an algorithm will resemble a curve, containing small waves as well as straight lines, and it will lie well above the random 45 degree line. When the performance curve of an algorithm hovers around the random line, it is an indication that more data is needed for training purposes (Janus & Misner, 2011, pg. 358).

Visible in lower right-hand corner of Fig. 6.6.1 is the mining legend, which can be used to determine the best probability threshold to apply in predictions (MacLennan et al., 2009, pg. 171).

The position of the graph's vertical gray line marker, which can be seen in Fig. 6.6.1's mining legend window, indicates that using 92.31% of the input population would capture 100% of the target (properly predicted by the model cases), whereas the 'ideal model' would capture 100% of the target (properly predicted by the model cases), using slightly more than 23%. Furthermore, DT_Model1 would only start to make predictions after having processed slightly more than 45% of the input data!

A description of the function of the mining legend from Microsoft's site ('Lift Chart [Analysis Services – Data Mining]') allows for interpreting the additional information contained in Fig. 6.6.1. Inspecting the mining legend, it can be seen that there is only a 6.25% chance that in the 100% of the captured by the model's target variable, at the vertical marker line, has a score of '9'.

Another value presented on the mining legend', the 'score,' is derived by calculating the effectiveness of the model across a normalized population, with a higher score being better ('Lift Chart [Analysis Services – Data Mining])'. 'The score associated with a mining model expresses the performance of the respective model as a fraction of the performance of the ideal model' (MacLennan et al., 2009, pg. 171)

The score of 0.30 for DT_Model1 is very low (from the author's professional experience, these scores, in the field, are typically above 0.70). However, this is not surprising, as, per the statement quoted in the previous paragraph, the vast majority of the performance model lies below the random line.

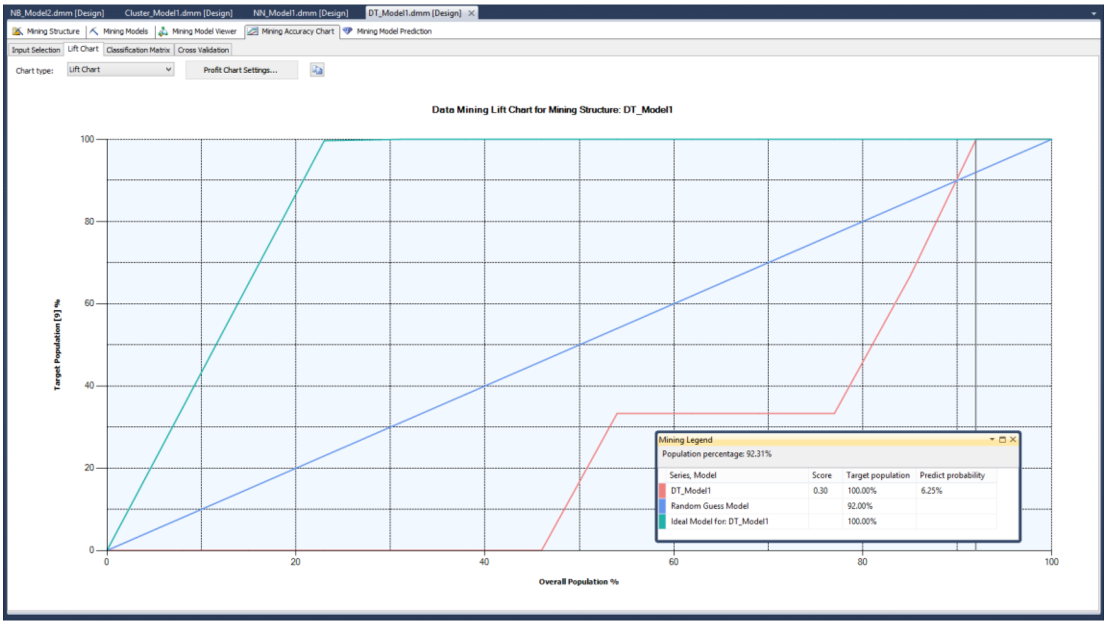


Chart type: Lift Chart

Population percentage: 92.31%

	Series, Model	Score	Target population	Predict probability
🟥	DT_Model1	0.30	100.00%	6.25%
🟦	Random Guess Model		92.00%	
🟩	Ideal Model for: DT_Model1		100.00%	

Fig. 6.6.1: Lift chart for DT_Model1 and input value '9'

The lift chart for NB_Model2 presented in Fig. 6.6.2 is very similar (which is rather unusual in business models, as there are usually large amounts of input data available for use in analysis and testing) to the lift chart presented in Fig. 6.6.1. Therefore, the discussion that took place with relation to Fig. 6.6.1 will not be repeated here. Interesting to note, however, is the score of the NB_Model2, which is equal to 0.50 and superior to the score of the DT_Model1. That is not surprising, as it can be visually seen that the performance curve of the NB_Model2, when compared to that of the DT_Model1, to a larger extent exceeds (lies above) the random line. In any event, the lack of an appropriate amount of data prevents a full investigation of the results and the quality of prediction beyond simply illustrating how to go about doing so.

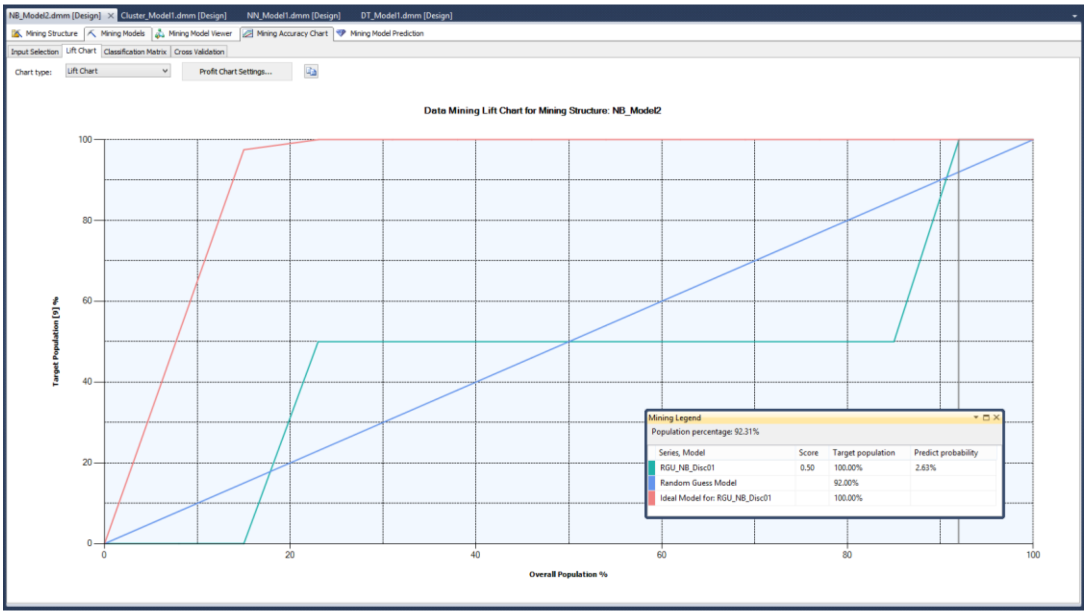


Chart type: Lift Chart

Population percentage: 92.31%

	Series, Model	Score	Target population	Predict probability
🟩	RGU_NB_Disc01	0.50	100.00%	2.63%
🟦	Random Guess Model		92.00%	
🟥	Ideal Model for: RGU_NB_Disc01		100.00%	

Fig. 6.6.2: Lift chart for NB_Model2 and value '9'

Scatter plots, which illustrate the predictive ability of models that work with continuous types of output variables, receive relative little attention in the practitioner's literature. Authors and practitioners such as de Ville (2001), Janus and Misner (2011) and Larson (2012) entirely omit the discussion of sctter plots. MacLennan et al. (2009, pg. 173) state that scatter plots compare actual values with those that were predicted: 'In a perfect model, each point would end up on a perfect 45-degree angle, indicating that the predicted values exactly matched the actual values.' The meaning of the 45-degree in the scatter plot graph is very different from that in the lift chart, where it represents a random guess. In a scatter plot, the 45-degree line indicates the 'perfect prediction,' as it is used to map actual values to those that were predicted.

The scatter plots presented in Figures 6.6.3 and 6.6.4 are also negatively affected by the very small amount of data. This negative effect is even more profound in the case of the graph addressing the accuracy of the Cluster_Model1, where there is a large number of variables (individual questions instead of the competence areas). In the case of Fig. 6.6.3, the values of output variables actually form a horizontal line rather than aligning themselves along the 45-degree line.

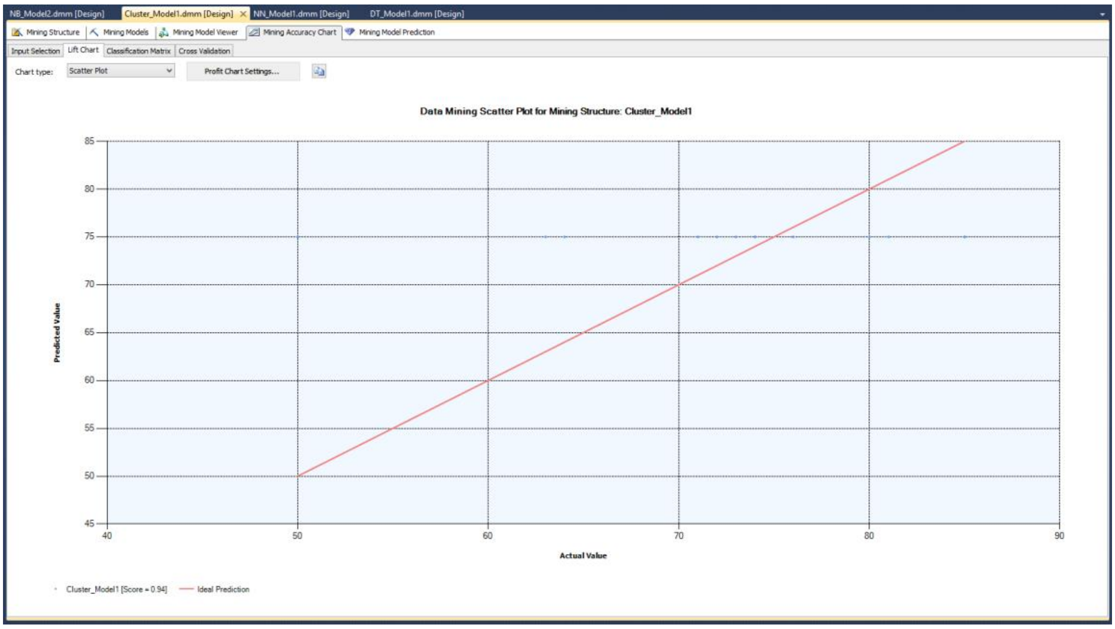


Chart type: Scatter Plot

Fig. 6.6.3: Scatter plot for model Cluster_Model1 (Note: hard to view data elements are aligned horizontally on the line labeled as 'Y = 75'.)

The output values of the NN_Model1, while still very far from acceptable, at least align along the 45-degree line somewhat. (It can be expected that, typically, the distance from the predicted output value to the 45-degree line is as large as or smaller than the point X = 74, Y = 73 in Fig. 6.6.4. The points near the X=80 are entirely off.)

Another tool for the evaluation of the qualities of predictions made by algorithms with non-continuous output variables is the classification matrix, a tool that is built-in to Microsoft's SQL Server 2012 (and other versions of the SQL product) platform. (The classification matrix cannot be built for continuous types of output variable.)

Larson (2012, pg. 666) states that '[w]e know our mining models are not going to do a perfect job of predicting. They are going to make mistakes. The Classification Matrix lets us see exactly what mistakes our models have made.' The errors made by models can be costly, especially considering the decision cost/benefit analysis discussed by Larose and Larose (2015, pg. 462). The costs of incorrect classification could result in a loss of a business or opportunities, depending on the type of error, be it a false positive or false negative, concepts discussed in Section 5.6 and Section 6.6.

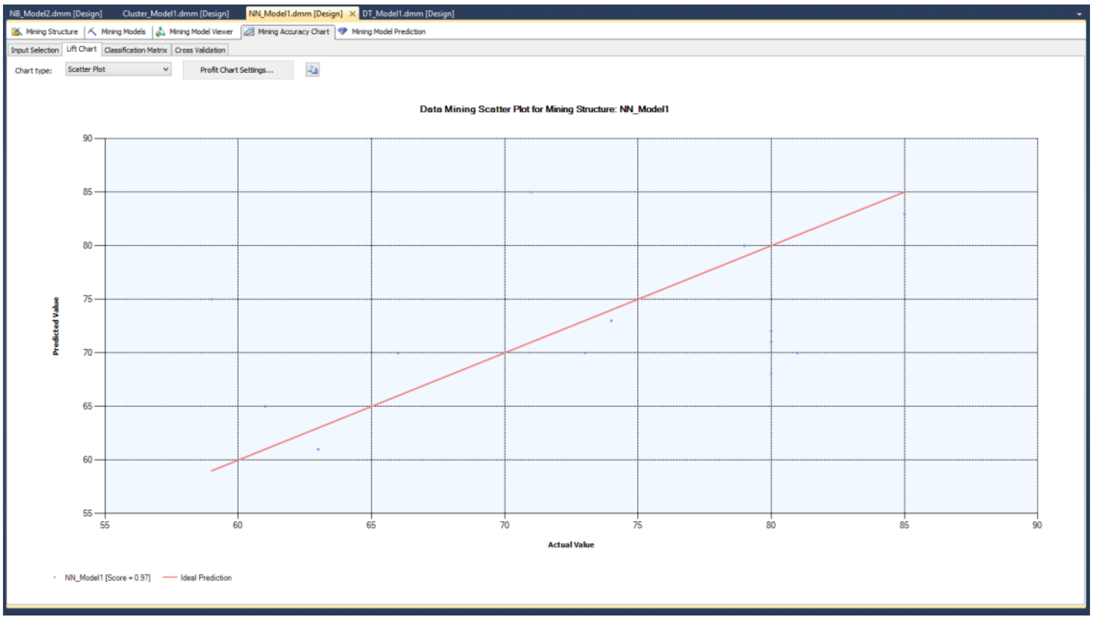
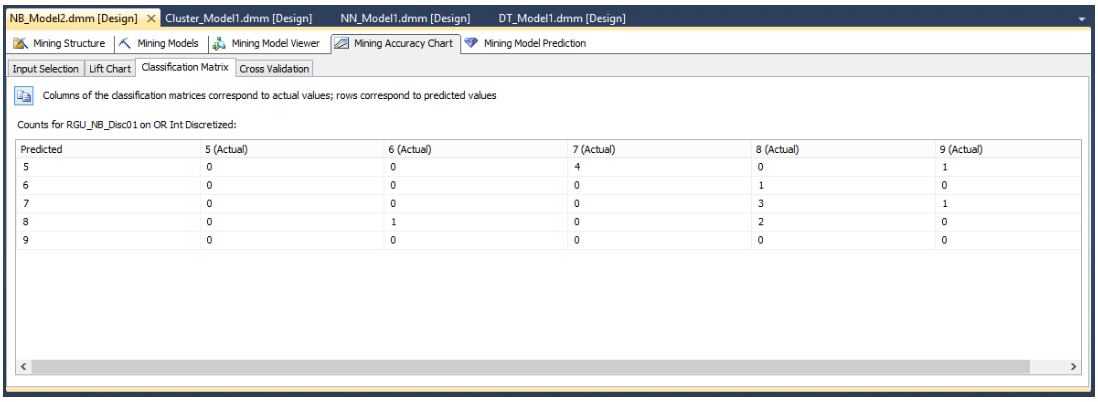


Chart type: Scatter Plot

Fig. 6.6.4: Scatter plot for NN_Model1

The classification matrix for the NB_Model2 is presented in Fig. 6.6.5, and it illustrates the results of prediction using the hold-out test data (30% of hold-out data was used in testing all of the models constructed in this work). As expected due to the very limited amount of data, the model did not perform well, confirming all prior indications of poor performance. In all predictions, the model was only able to predict the outcome of '8' twice, which still constituted only two thirds of all predictions. It is expected that the counts of the correct predictions would be high, with only the occasional incorrect classification.



Columns of the classification matrices correspond to actual values; rows correspond to predicted values

Counts for RGU_NB_Disc01 on OR Int Discretized:

Predicted	5 (Actual)	6 (Actual)	7 (Actual)	8 (Actual)	9 (Actual)
5	0	0	4	0	1
6	0	0	0	1	0
7	0	0	0	3	1
8	0	1	0	2	0
9	0	0	0	0	0

Fig. 6.6.5: Classification matrix for model: NB_Model2

The final tool for evaluation available as a part of Microsoft's SQL Server is cross-validation.

‘Cross-validation is a sampling technique used primarily for small data sets, when data is too small to partition into training and testing subsets’ (Abbott, 2014, pg. 130).

The cross-validation technique, also referred to as ‘k-fold,’ is described by Abbott (2014, pg. 131) as involving the following three steps:

1. The creation of k distinct sets (folds) of data using random technique. (In this work, the random split into k-sets, or folds, is also performed by the cross-validation tool; however, the number of folds is specified as one of the parameters of the program);
2. From the k-folds, one fold should be designated for testing and k-1 for training. Begin by using subset 1 for testing and the remaining (2-k) subsets for training; and
3. The roles should be rotated so that, at the end, each subset is used once for testing and k-1 for training.

Some technical points that affect the operation and outcomes of cross-validation mentioned by Abbott (2014, pg. 131) include the following:

- The larger the number of folds used, the smaller the hold-out testing subset and more error variance will be observed;
- The average error over each fold is more important than an error for each fold; and
- Can be used to assess model stability. ‘The model accuracy on the k testing subsets can be used to assess how stable the models are: If the accuracy is similar for all folds, the modeling procedure is viewed as being stable and not overfit.’ The average error over each fold can be used to estimate a model’s stability.

MacLennan et al. (2009, pg. 174) also identify some additional points concerning cross-validation, some of which are specific to the Microsoft-platform:

- The cross-validation technique can be used to determine how suitable the data is for model training;
- Given the input data, the cross-validation technique can be used to determine which algorithm is best suited for modeling, without actually building the models;

- The types of accuracy measurement in the SQL Server depend on the type of the algorithm used by the model being evaluated (for example, clustering measurements are different from classification or regression) and the type (discrete/continuous) of the output variable;
- The results of the cross-validation tool need to be checked (at each partition model) in two ways: 1) How accurate the results are – if all of the model's partitions have good accuracy results, using the full data set should also lead to good results, and, 2), if all partitions/folds show similar results, the training set is appropriate for the current DM task. (Differences suggest that partitions have significantly different data distributions);
- The default number of folds in SQL Server 2012 is ten;
- Setting the 'Max Cases' parameter to zero, the default value, will cause all DM training data to be used in cross-validation; and
- Setting the optional 'Target State/Target Value' (depending on the algorithm type) to a valid state or value will test how well the model/s predict/s the output value of the 'Target Attribute'. Leaving the 'Target State' empty will determine the overall accuracy of the model/s.

The online documentation provided by the Microsoft Corporation (Cross-Validation – Analysis Services – Data Mining) adds important information about setting the accuracy threshold needed for the generation of the cross-validation tests. When the value of the parameter 'Target Threshold' is NULL, the predicted state with the highest probability is considered the target value. Otherwise, the value of the field can take on values between 0.0 and 1.0, where numbers close to 1 indicate a strong level of confidence in the predictions and numbers close to 0 indicate that the prediction is less likely to be true. The value of the 'Target Threshold' affects the measurement of a model's accuracy. (Setting the value to 0.0 will make every prediction count as correct.)

Figures 6.6.6 and 6.6.7 illustrate the use of the cross-validation technique; Fig. 6.6.6 presents an output of the cross-validation when no target state has been selected.

With three folds and 32 elements in the input data set, the folds hold between 10 and 11 elements, as was the case in the previous discussion.

The first accuracy measure, ‘Classification – Pass’, with values 8, 9, 9, indicates how many correct classifications of the target attribute ‘Is OR’ were performed (considering the value of the ‘Target Threshold’ parameter).

The standard deviation of 0.475 indicates that the partitions differ by about 5.5%, indicating that the results appear to be reasonably compact.

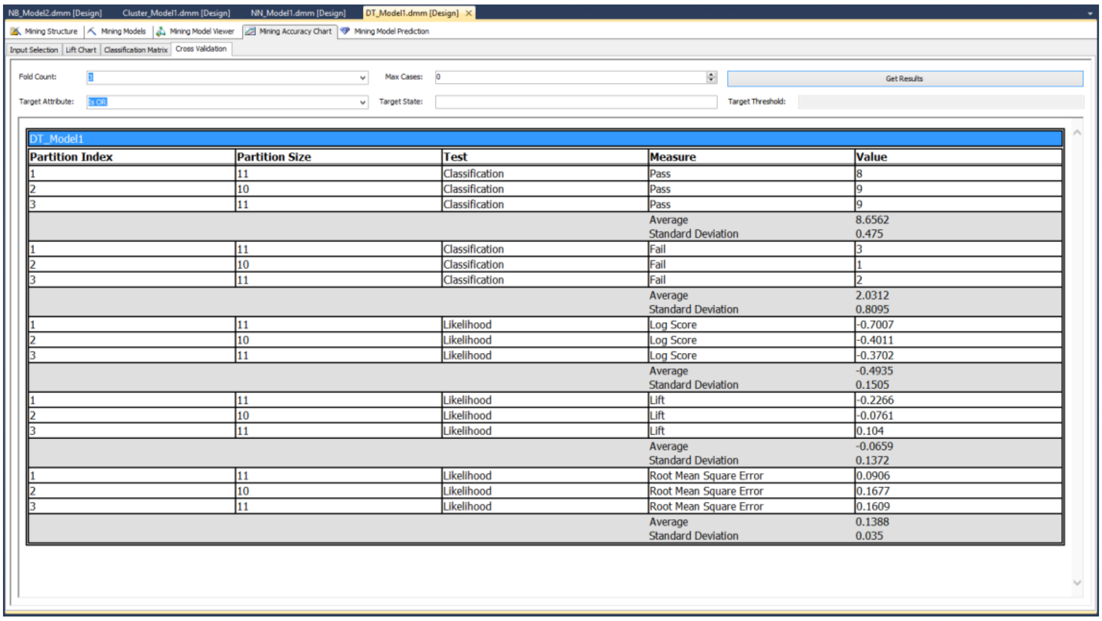
The second set of measurements, ‘Classification – Fail’, provides information about the number of incorrect classifications of the target that were encountered during the evaluation. The numbers vary between 1 and 3. With a standard deviation of 0.8095, or nearly 40%, the results do not look encouraging for real-life analysis, as they vary widely.

The log score, always negative, of the Likelihood test, according to the on-line documentation (Cross-Validation [SQL Server Data Mining Add-ins]), represents the ratio between two probabilities, converted into a logarithmic scale, with a log score closer to 0 being better. The likelihood for the prediction is rather poor, as it is not close to zero.

The ‘Root Mean Square Error’ measure, as defined in the on-line document (Cross-Validation [SQL Server Data Mining Add-ins]), ‘is the average error of the predicted value to the actual value.’

Fig. 6.6.7 illustrates the cross-validation for the same model (DT_Model1) as Fig. 6.6.6, with the only difference being that, this time, the target state (equal to ‘True’) has been specified.

The primary difference between the two outputs from the cross-validation method is that, when there is target state specified, four classification tests are carried out, in addition to the likelihood tests. As shown in Fig. 6.6.7, the classification test for measure ‘True Positive’ returned all zeroes (no hits for True Positive’). The ‘False Positive’ test returned one in one of the folds and zero in the other two. It also returned, proportionally to the average, a very large standard deviation.



Fold Count: 3
Max Cases: 0
Target Attribute: Is OR
Target State:
Target Threshold:

DT_Model1

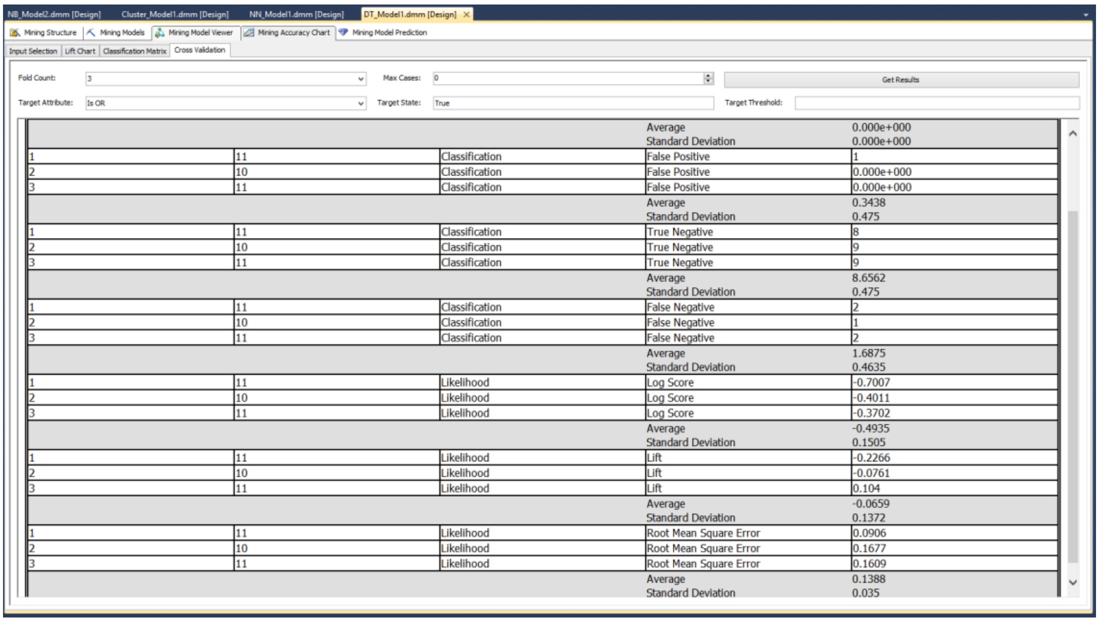
Partition Index	Partition Size	Test	Measure	Value
1	11	Classification	Pass	8
2	10	Classification	Pass	9
3	11	Classification	Pass	9
			Average Standard Deviation	8.6562 0.475
1	11	Classification	Fail	3
2	10	Classification	Fail	1
3	11	Classification	Fail	2
			Average Standard Deviation	2.0312 0.8095
1	11	Likelihood	Log Score	-0.7007
2	10	Likelihood	Log Score	-0.4011
3	11	Likelihood	Log Score	-0.3702
			Average Standard Deviation	-0.4935 0.1505
1	11	Likelihood	Lift	-0.2266
2	10	Likelihood	Lift	-0.0761
3	11	Likelihood	Lift	0.104
			Average Standard Deviation	-0.0659 0.1372
1	11	Likelihood	Root Mean Square Error	0.0906
2	10	Likelihood	Root Mean Square Error	0.1677
3	11	Likelihood	Root Mean Square Error	0.1609
			Average Standard Deviation	0.1388 0.035

Fig. 6.6.6: Cross-validation for the DT_Model1, 'Is OR' with the number of folds = 3 and with no target state specified

There were between 8 and 9 (average of 8.66) cases of 'True Negative' measures identified in three folds, with the standard deviation of 0.4635 (or about 5%) meaning that the results were rather compact.

The 'False Negative' measure resulted in counts between 1 and 2, with an average of 1.6875. The standard deviation of 0.4635, or 27%, indicates quite spread in numbers. The tests for the likelihood presented in Fig. 6.6.6 still hold true for Fig.6.6.7.

As a final comment regarding the validation of the DM model, it is important to emphasize the role of domain experts in evaluating the outcomes. From the professional experience of the author of this work, many times it can be seen in the field that the results of a DM model are accepted without any scrutiny by the individuals who use the knowledge generated by these models. Yet, some time later, the flaws of the model become visible when someone accidentally discovers the discrepancy between the output of DM models and common sense. The need for the use of domain experts has been emphasized recently by the extension of the DM field into DDDM (domain-driven data mining). The concept of DDDM was introduced in Section 2.5.



Fold Count: 3
Max Cases: 0
Target Attribute: Is OR
Target State: True
Target Threshold:

			Average Standard Deviation	0.000e+000 0.000e+000
1	11	Classification	False Positive	1
2	10	Classification	False Positive	0.000e+000
3	11	Classification	False Positive	0.000e+000
			Average Standard Deviation	0.3438 0.475
1	11	Classification	True Negative	8
2	10	Classification	True Negative	9
3	11	Classification	True Negative	9
			Average Standard Deviation	8.6562 0.475
1	11	Classification	False Negative	2
2	10	Classification	False Negative	1
3	11	Classification	False Negative	2
			Average Standard Deviation	1.6875 0.4635
1	11	Likelihood	Log Score	-0.7007
2	10	Likelihood	Log Score	-0.4011
3	11	Likelihood	Log Score	-0.3702
			Average Standard Deviation	-0.4935 0.1505
1	11	Likelihood	Lift	-0.2266
2	10	Likelihood	Lift	-0.0761
3	11	Likelihood	Lift	0.104
			Average Standard Deviation	-0.0659 0.1372
1	11	Likelihood	Root Mean Square Error	0.0906
2	10	Likelihood	Root Mean Square Error	0.1677
3	11	Likelihood	Root Mean Square Error	0.1609
			Average Standard Deviation	0.1388 0.035

Fig. 6.6.7: Cross-validation for the DT_Model1, 'Is OR' target variable and target state = true and the number of folds = 3

6.7 Discussion of Findings

The modeling sections were the main focus of this chapter, as they attempted to answer DM-related research questions and to fulfill the aims of this research. Due to the unusual positioning of this research (as research of an applied type being conducted for a professional degree and, rather than obtaining a numerical outcome from the applied research, it presents an evaluation method), the following sections serve to both continue the presentation of findings and to discuss them.

While examining the findings related to DM in the context of the research questions and the aim of this thesis, the following sections provide the answers to research questions #3, #4 and #5, as well as to the central purpose of this research. (Research questions #1 and #2 were addressed in the literature review in Chapter 3.) To help formulate the findings, the tags (RQ #3) and (RQ #5) are used in the discussion of the DM models. Informed by the literature review, the construction of the described earlier in this section models provided the following answers:

6.7.1 Findings: RQ #3

In Chapter 1 and Chapter 4, research question #3 was identified as follows:

> Which KM processes are the most influential for OR?

The objective for asking RQ #3 was exploring the use of DM in order to test the suitability of applying DM to the primary grouped data, which was comprised of the questionnaire answers, to assess their relationship with OR.

When searching for an answer to RQ #3, two approaches were initially considered. The first approach was to investigate individual KM activities (represented by a single questionnaire's question) and these activities' impact on OR. The second approach was to investigate the impact on OR of KM activities grouped into competence areas (discussed in Sections 3.2.5 and 4.6.2).

The literature review completed in Chapter 3 rarely mentioned a single KM activity that impacted businesses performance or resilience. Instead, the writers discussed the impacts of KM processes on performance, either directly (McKenzie & van Winkelen (2004), Green (2006), Cool & Zhan (2006), Brusilovski & Brusilovski (2008), Ngai et al. (2009), Moayer & Gardner (2012) and Fuchs et al. (2014)) or indirectly (Lee (2008), Adejuwon & Mosavi (2010), Li et al. (2012), Natek & Zwilling (2014) and Chemchem & Drias (2015)). In addition to the limited literature coverage of the impact of individual KM activities on OR, there are other reasons for considering KM processes when using DM as a tool for analysis. That is not to say, however, that no KM activities were discussed when investigating the impact of KM on OR/OP. The questions developed for this research were derived from the work of McKenzie and van Winkelen (2004); as such, the individual KM activities were discussed in great depth, yet the KM activities were grouped by the authors into competences when discussing how they impacted OR/OP. The literature review chapter material did not discover any work that specifically discussed the impact and measurement of the impact of a single, well-defined KM activity on OR/OP.

When this research attempted to analyze the impact of individual KM activities (represented as a single questionnaire's question) on OR, several issues were encountered.

Some of the issues were associated with the management of the large number of input variables, as can be seen in in Appendix VIII, figures A8.15 – A8.20 and A8.22. The more significant problem, which perhaps was magnified because of the small input data set used in the research, as, typically, the larger the number of variables, the larger the input data set required (Andonie, 2010), related to the interpretation of the results. As pictured in A8.47 in Appendix VIII, examining a network diagram that contains 52 variables is difficult. Similar difficulties in interpretation would be encountered when inspecting the composition of the 52 clusters, for example. For this reason, the grouping of the input variables was chosen for this research, which is a common industry practice (referred to by Han et al. [2012, pg. 85] as dimensionality reduction). Should there be enough input data available, the analysis of the individual KM activities could be considered using all of, or a subset of, responses, such as the subset of responses contained within the competence area with the highest 'OR Score'. Should the individual KM activities be involved in the investigation of their impact on OR, the findings of this research are fully applicable to that scenario as well. The difference would be purely in the granularity of the data.

Based on the groupings of the data into McKenzie and van Winkelen's (2004) competence area framework, the following are the findings that address RQ #3, on a per DM model basis.

NB_Model1, described in the classification model in Chapter 6.2, was constructed to illustrate common aspects of DM development environment, to 'get a feel' for the data and to detect various variable relationships; as such, it is not discussed further.

NB_Model2, described in the classification model in Chapter 6.2, allowed for the determination of relationships (through the functionality called a dependency network) between six competence areas and one output variable, 'OR Int Discretized,' along with the determination of the strongest relationships. The dependency network showed that three input variables ('Link', 'Create' and 'Learn') predicted the output variable 'OR Int Discretized' (holding values 5 – 9, where the higher the number, the more resilient the organization), with the link from the 'Create' variable being the strongest. The knowledge about the strongest correlation can help an organization to improve (or focus attention on) the KM processes that impact OR the most, which can support a number of works: according to Law and Ngai (2008), this information can be used to examine the relationship between knowledge sharing and learning behaviors and business performance, while Handzic (2009) claims that it can help in understanding the value offered by KM to an organization. The other area where DM proved to be highly effective and useful was the analysis of the attribute profiles that allowed checking the composition of each cluster in terms of the KM processes present when a specific value of the output variable (OR) was chosen, thereby validating and providing additional insights into Lina and Tsen's (2005) KM implementation gap's impact on OP/OR, Braes and Brooks' (2010) identification of the essential KM processes that must be present in resilient organization and Lee's (2008) focus on four KM processes. The probability of the specific value impacting the competence area on the OR was provided through the 'attribute characteristics,' while 'attribute discrimination' allowed the determination of what made any two competence areas different with respect to each other (or with respect to all other competence areas) given specific OR values, which was expressed in terms of the probability percentage. Finally, the predictive ability of the NB algorithm allowed for arriving at the 'OR Score' (the resilience score) for a given set of answers to the questionnaire, taking the work of Brusilovski and Brusilovski (2008), Kowlaczyk et al. (2013), Cot-Real et al. (2014), Natek and Zwilling (2014) and Hopkins and Schadler (2015) one step further.

NN_Model1 – The neural network model was presented in the context of six competence areas, designated as the continuous type of input variable, and the 'OR Integer,' also treated as the continuous type of output variable. Because of the complexities of the NN algorithm (MacLennan et al. (2009, pg. 382), the tools that were part of other algorithms were not available for the NN model; therefore, the discussion of the model concentrated on the model's parameters and the composition and values of the input attributes that resulted in a given range (being a continuous type) of values of the output attribute. The prediction using the NN model resulted in an 'OR Score' value of 68, when the input data common to all models was supplied to NN_Model1 for prediction.

The NN classification model allowed for an investigation of the probabilities of favoring certain competence areas when considering two unique values, or 'OR

Scores,' of the OR. While such inspection was also available using NB_Model1 and NB_Model2, the unique insight offered by the NN-based algorithm has to do with the fact that the NN-based algorithm allows selection of one or more competence areas (input variables), along with the values attained by them, and seeing how such selection impacts OR. Such configuration allows conducting simulation scenarios and seeing the impact of the combination of competence area and value of the competence area on the selected OR output levels. The prediction capability of the NN model allows for arriving at an 'OR Score' based on the questionnaire answers entered. The model's benefits, in turn of associations with the existing literature, are the same as those identified in the discussion of NB_Model2.

DT_Model1 was a model, based on the decision trees algorithm that used six competence areas, with each competence area designated as a 'discrete' type. Two output variables were used, 'Is OR' or the Boolean type (yes/no), to predict if the data supplied for prediction would result in an outcome of 'resilient organization' (a resilient organization was an organization that reached some predetermined number of points based on the answers it gave); 'OR Int Discretized', a discrete integer, was used to provide a predicted 'OR Score'. The classification model constructed in Chapter 6.5 produced three main outputs. The first result was a constructed tree that allowed determining how each competence area impacts OR. The tree constructed showed the key competence area in terms of influencing OR at the selected 'OR Score' level, along with the value competence area needs to achieve in order to be classified as such. The second useful outcome, similar to the dependency network of NB_Model2, was the construction of the network diagram that showed the competence areas and their strength (as the strongest links) in predicting OR.

The dependency network generated for the DT model showed that the 'Learn', 'Create', 'Decide' and 'Connect' competences predicted the 'OR Discretized' node and that the 'Decide' and 'Connect' competences predicted the 'Is OR' output variable. In addition to the dependency network, two decision trees, one per each output variable ('IsOR' and 'OR Int Discretized'), were created and discussed. The critical nature of the input parameters (Provost & Fawcett (2013, pg. 81) and MacLennan et al. (2009, pg. 256)) was illustrated by recreating the DT model, using different scenarios involving the input parameters. Finally, the prediction model, when supplied with the standard input data used for prediction in all models, returned 7 for the value of the 'OR Int Discretized' output variable and 'False' for the 'Is OR' output variable.

Finally, the model allowed the prediction of two types: it allowed predicting if a given set of questionnaire replies represents whether or not an organization is resilient (with the answer being true/false), and, based on the same data, it allowed arriving at the 'OR score'. The easily interpretable 'if-then-else'-style results (Larson, 2012, pg. 611) make the model a natural choice for both researchers and practitioners who seek to determine the impact of KM on business through improved decision-making, such as Shollo and Kautz (2010) and Kowalczyk et al. (2013).

It needs to be pointed out that the construction of the DM models did not have to use the McKenzie and van Winkelen (2004) framework or the KM processes model proposed by Burnett (2004, pg. 29). Instead, one can rely on the DM clustering algorithm to segment the questionnaire replies (KM activities) into clusters where the similarity of the questions in a cluster is maximized and similarity outside of the cluster is minimized. This highlights the flexibility of the approach, which allows for future theoretical developments being incorporated into further iterations of the model.

Cluster_Model1, described in the Section 6.3 clustering model provided method for segmentation of KM activities (as opposed to KM processes) into relatively homogeneous subgroups with the 'Clustering Diagram,' allowing for identification of the desired level of OR with respect to the input KM activities and the display of the level of correlation between resultant clusters. 'Cluster Profiles', 'Cluster Characteristics' and 'Cluster Discrimination' served functions similar to those described above for NB_Model2 except that, instead of attributes, the model investigated the composition of clusters. Such categorization allows greater understanding of which KM activities, and to what extent (in terms of probability percentage), form groups with certain levels of OR. Cluster characteristics and discrimination with respect to the percentage points of probability of an outcome occurring allowed investigation of a given KM activity within a selected cluster and investigation of the KM activities that distinguished between two clusters or between one cluster and all of the remaining clusters, respectively. Microsoft's specific implementation of the clustering algorithm allowed for prediction of a resultant cluster given a set of questionnaire answers. Cluster_Model1 is very well suited to enhance the highly specific DM approach used by Leung and Joseph (2014) for comparing the composition of sport teams and to advance the research of Natek and Zwilling (2014) by providing more sophisticated tools for student segmentation and the inspection of student segments, to mention but a few of its promising applications with respect to the reviewed literature.

The findings presented in Chapter 6 demonstrated that DM is an excellent tool for determining which processes have the greatest impact on OR. An additional discussion of the suitability of DM for determining the impact of KM on OR takes place in Chapter 7.

6.7.2 Findings: RQ #4

> Can a methodological approach be developed to examine the relationships between KM and OR, utilizing DM?

The objective for RQ# 4 was stated as 'to develop and apply a DM-based methodological approach in relation to the analysis of data gathered from the use of the questionnaire instrument and the generation of valid findings for this research.'

Using the five distinct DM models created for the purpose of this research, it can be stated that it is possible to develop a methodological approach for examining KM's impact on OR using DM.

Based on this research performed for this work, the methodology for employing DM in the above-mentioned scenario could be stated as requiring the steps presented in Table 6.7.2.1, below. (The activities are listed in the table in order of occurrence, but they the process may involve looping back to earlier steps; these loops are not shown in the tabular representation. Indentation implies that a task is a sub-task of the not indented task immediately above it.)

Activity:	Outcome/Reason:
Understanding the business problem.	To identify a clear objective for the DM-based project, as well as

	to create a project plan.
Designing the data collection instrument.	Creation of the data collection instrument.
Validating the data collection instrument.	To assure the instrument measures what it is intended to.
Distributing the data collection instrument.	Ensure that the selected sample or the entire population receives the questionnaire to be completed.
Collecting data.	To collect responses to the data collection instrument.
Understanding the data collected.	To identify the DM task to use with the data.
Analyzing the data.	To arrive at statistics about the data collected and to identify outliers.
Ensuring the quality of the data.	To determine if the collected data is suitable for analysis.
Auditing the data.	To identify any problems with the data, examine data trends and compute summary statistics.
Preparing the data.	To assure the data to be used in analysis by the DM is free of major problems (or that such problems, if present, have been properly addressed).
Cleaning the data.	To correct any possible data issues.
Transforming the data.	To perform any necessary data transformations (per the requirements of the DM used).
Modeling	To create DM models.
Selecting a model.	Selection of the most suitable DM model based on the

	understanding of the business problem and of the data collected.
Evaluating the model.	To further assist in selection of the most appropriate model for the problem.
Analysis	Results of the DM modeling are analyzed.
Application	The outcome of the DM modeling is applied to the real-life situation impacting OR.

Table 6.7.2.1: Methodological approach to data mining

While the methodology for conducting DM modeling presented in Table 6.7.2.1 provides an effective way of organizing DM-based projects, it closely follows the industry standard CRISP-DM methodology. The CRISP-DM framework is widely used in commercial projects due to the wide coverage it receives in practitioners' publications as the framework of choice for implementing DM-based projects: LeBlanc et al. (2015, pg. 177), Abbott (2014, pg. 19), Larose & Larose (2014, pg. 4), Provost and Fawcett (2013, pg. 14), Janus and Misner (2011, pg. 350), MacLennan et al. (2009, pg. 86), Turban et al. (2007, pg. 327) and de Ville (2001, pg. 37). This research further contributes to the understanding of the applicability of the CRISP-DM framework and its potential for use in relation to OR.

6.7.3 Findings: RQ #5

> Which are some of the main challenges when employing DM for the purpose of determining the impact of KM on OR?

The objective of RQ #5 was stated as follows: to identify the main issues (data, algorithm, error and/or algorithm parameters) associated with the use of DM for the purpose of measuring the impact of KM on OR.

As has been illustrated in Chapters 3 and 6, working with DM algorithms is highly rewarding yet challenging. Each phase of building a DM model presents its own challenge. The issues related to RQ #5 are identified in Chapter 6, as well as in the rest of this thesis with the tag 'RQ #5' and are discussed next.

In general, the issues related to the DM modeling can be classified into three general categories: those that relate to the input data, to the output data and to DM.

Some of the key issues, along with references to the discussion in this work, are summarized in Table 6.7.3.1, with detailed discussions having been provided in Sections 5.3, 5.4, 5.6 and Chapter 6.

DM-related issue/obstacle (source):	**Category:**	**Possible associated risks:**
Data preparation: lengthy and tedious process requiring great database skills. (Abbott, 2014, pg. 83)	Data – input. Time and people management.	Incorrect, incomplete or incorrectly formatted data can lead to incorrect resulting DM models. (Discussed in the following Sections: 5.3, 5.4.)
Data transformation: standardize scales of numeric variables. (Larose & Larose, 2015, pg. 8)	Data – input, DM.	Non-uniform scales can give more weight to variables using larges scales (Sections 5.4.)
Data type: assure the data type is suitable for the algorithm chosen. (Han et al., 2012, pg. 84, MacLennan et al., 2009, pg. 174)	Data – input.	No DM model will be created. (Sections 5.3, 5.4.)
Data quality: missing or incorrectly entered data. (Larose & Larose, 2015, pg. 20, Han et al. 2012, pg. 85, Witten et al. 2011, pg. 60)	Data – input.	Unusable data set. (Section 5.3.2.)
Data size: appropriate	Data – input.	Unpredictable results

for chosen algorithm sample size. (Abbott, 2014, pg. 131, Andonie, 2010, pg. 280)		obtained as an outcome. (Sections 5.4.)
Results: making sense of the results. (Provost & Fawcett, 2013, pg. 31, Larson, 2012, pg. 666, Janus & Misner, 2011, pg. 357, Witten et al., 2011, pg. 60, MacLennan et al., 2009, pg. 175)	Data – output.	Misinterpretation of results or lack of context for the findings leading to no interpretation at all. (Section 5.6, 6.6.)
Parameter selection: setting algorithm parameter values to the correct values. (MacLennan et al., 2009, pg. 233,256,314,396)	DM	Very wide range of risks: from failure to create DM model to incorrect results. (Sections 6.2, 6.3, 6.4, 6.5, 6.6.),
Outcome repeatability: need to assure that the stability of the DM algorithm has been achieved. (Larose & Larose, 2015, pg. 319, Abbott, 2014, pg. 229)	DM	Unreliable algorithm, leading to incorrect results. (Section 5.6, 6.6.)

Table: 6.7.3.1: Important issues and obstacles for DM-based projects

As can be seen from the some of the possible problems and obstacles listed in Table 6.7.3.1, the use of DM for the purpose of investigating the impact of KM on OR must be guided by a carefully considered plan. One such possible framework is to follow is the CRISP-DM framework that has been successfully used in this research (discussed in Chapter 5 and Section 6.7.2).

Another very important aspect to note is the nature of the problems. While some of the issues will manifested and/or be recognizable and will stop the modeler from progressing further, other types of issues will not prevent the creation of the DM model but will simply result in the creation of a flawed model. To avoid these and other issues, the resulting DM models need to be evaluated as it was discussed in Chapter 6 (when describing each individual model) and in Sections 5.6 and 6.6.

Finally, when considering the specific algorithms used in this research, the following discussion represents the key issues encountered during the creation of the models:

NB_Model2 – Presented in Section 6.2, the non-categorical inputs produced no model. The dependent variable had to be normalized and made to be of discrete type to be usable (Abbott 2014, pg. 84). Too little data made the model of limited usability and model evaluation could only be carried out based on theoretical grounds. Setting the MINIMUM_DEPENDENCY_PROBABILITY too low may make the model insignificant or may result in incorrect interpretation (MacLennan et al., 2009, pg. 234).

Cluster_Model1 – Input data with a random structure can lead to an inaccurate model. Specifying the optimal number of clusters is highly technically and mathematically involved, unless the data fits some natural groupings or the algorithm is allowed to arrive at the optimal number of clusters (Abbott, 2012, pg. 185; Han et al., 2012, pg. 484). Data skew should be reduced whenever possible and its distribution normalized in order to receive appropriate DM results from a clustering algorithm (Abbott, 2014, pg. 183, Han et al., 2012, pg. 47). Categorical variable are, generally, not to be used (Abbott, 2014, pg. 183). Because of the recursive nature of the inner workings of the clustering algorithm, processing can be very taxing to the computing environment, unless the scalable framework of MS SQL Server is used (MacLennan et al., 2009, pg. 314). Some of the clustering methods may produce no DM model if an incorrect value is used for CLUSTERING_METHOD / MINIMUM_SUPPORT (MacLennan et al., 2009, 314).

NN_Model1 – The hidden layer makes it impossible to follow the execution of the algorithm, leaving the modeling entirely to the tool; this is particularly true if the NN-based model is of the unsupervised type. The parameter values for the NN-based algorithm must be determined empirically. Finally, the NN-based DM algorithm is not well suited to small (input) data sets (MacLennan et al., 2009, pg. 396; Andonie 2010, pg. 280).

DT_Model1 – To deal with the model's stability, the DT-based algorithm has a single chance to build the model correctly or, more precisely, to select the appropriate internal nodes (Abbott, 2014, pg. 229). As such, the model needs to be re-constructed several times, after removing the variable originally placed at the root. It may be necessary to consider more than one DT-based model as the final solution. The preference for high cardinality data can be a model accuracy issue for the resulting model, and the target variable must be of the discrete type. Typically, the resulting model, due to the single chance to build an optimal model and the non-loopback learning style, can have a tendency to not be as accurate as other models (de Ville, 2001, pg. 78; MacLennan et al., 2009, pg. 247); however, this can be overcome using the methods outlined in Section 6.5.

6.7.4 Findings: Research Aim

> Aim of research: to test the feasibility of using DM to assess the relationship between and impact of KM on OR.

In discussing the extent to which the research aim has been realized, a number of aspects must be considered. As, clearly, not all of the aspects associated with meeting the research aim are discussed in this section of this thesis, a discussion of the omitted issues with relation to the aims of the research is provided in Chapter 7, which focuses on the conclusions reached.

Based on the prior sections of this work, the answer to the central purpose of this research must, at a minimum, consider the following factors: the preference for DM methods over traditional statistics, the DM models constructed, the CRISP-DM framework that underlies the DM project and the DM tool itself. The remaining sections focus on these factors.

6.7.4.1 DM Methods

From the professional perspective and from the summary of the literature review, as well as based on the discussion in Section 4.9, DM models have, when properly constructed, an ability, unmatched by that of classical statistics, to discover intricate, non-linear relationships between many variables at once (Gullo (2015); Fuchs et al. (2014), Moyar and Gardner (2012); Brusilovski and Brusilovski (2008)). Based on the findings presented in this thesis regarding the relationship between KM and OR, DM has been proven to be a highly viable instrument for the measurement of the impact of KM on OR (with OR being defined as in Chapter 3.3 of this study).

6.7.4.2 DM Models

The models presented in Chapter 6 and the findings of Section 6.7 provide practical information regarding how DM models can be used to evaluate the impact of KM on OR; this represents a contribution to the body of work that addresses the ability of measuring the impact of KM on OR using DM. As of the time of writing, the two works identified that directly addressed the impact of KM on OR differ significantly from this thesis, despite the fact that they also used DM as a measurement instrument. The work of Choi et al. (2008), while measuring organizational performance, used a KM strategy, not KM processes, as the independent variables. The other work identified, that of Wu et al. (2010), used KM processes as independent variables but uses ROA, instead of OR, as the measured variable. In addition to helping to understand the impact of KM on OR, DM allows performing various analyses of the composition of answers at various levels (various ranges of OR points) of OR and the probabilities of finding answers at any OR level.

While there were some challenges related to the building of the DM models (those challenges were described in detail in Chapters 5 and 6 and were summarized in the previous section), in general, given an appropriate amount of input data of a satisfactory quality, one can expect a “workable” number of issues to arise in the construction of a DM model (Larose & Larose, 2015, pg. 8; Han et al., 2012, pg. 84; MacLennan et al., 2009, pg. 174); the solutions to the most common and important issues were presented in Chapters 5 and 6.

Not every possible model was constructed in this study. The choice of models and the justification for their selection was discussed in Section 5.5.2. The DM models constructed, with the exception of the NB models due to the nature of the algorithm used (Larson, 2012, pg. 613; MacLeennan et al., 2009, pg. 217), consider multiple variables and their effects on each other and the dependent variable simultaneously. This, in itself, represents a clear advantage over the hard-to-compute solutions that use traditional statistics; this could be a significant contribution to the work of McCann et al. (2009), which manually computed various probabilities and created the attribute dependency graph for one model (2009, pg. 9). As was illustrated in Chapter 6, DM models make it possible to discover relationships, such as those between KM and OR, that are difficult to detect otherwise.

Per the discussion in Sections 5.5, 5.6 and Chapter 6, certain types of DM algorithms perform better when resolving certain types of problems. For this reason, the performance of each algorithm should be carefully evaluated and contrasted with the results of other models, employing domain experts for interpretation of the results when possible (Adejuwon & Mosavi (2010), Shih et al. (2010), Shollo & Galliers (2013)).

6.7.4.3 CRISP-DM Framework

In order for this research's DM project to be completed with a high degree of success (considering the model's accuracy as a measure of success), the project followed the industry standard CRISP-DM model, which has been embraced by a number of writers, including LeBlanc et al. (2015, pg. 177), Abbott (2014, pg. 19), Larose and Larose (2014, pg. 4), Provost and Fawcett (2013, pg. 14), Janus and Misner (2011, pg. 350), MacLennan et al. (2009, pg. 86), Turban et al. (2007, pg. 327), and de Ville (2001, pg. 37). Each stage of the CRISP-DM framework was presented in Chapter 5, further contributing to the practical understanding of the applicability of the CRISP-DM framework and its potential for use in relation to OR.

6.7.4.4 DM Tool

Finally, the DM tool selected provides the specific functionality required to investigate the impact of KM on OR. As was the case in the of work of Natek and Zwilling (2014), this research used Microsoft's technology to build the DM models and generate insights based on those models. Per Natek and Zwilling (2014, pg. 6402), this research uses the most sophisticated tools available from Microsoft, and the Microsoft platform, as presented in Section 4.9, is the world-leading analytical platform (Gartner, 2016).

6.7.4.5 Summary

The problems related to the relatively low applicability of problems solved by DM for business organizations have been listed by Hopkins and Schadler (2015, pg. 10) as one of the key problems encountered when turning data into actions: '[p]oor linkage between insights discovery and business action and scarce learnings from actions taken'. The views of Hopkins and Schadler have been shared by many recent writers (Cao & Zhang (2006), Brusilovski & Brusilovski (2008), Adejuwon & Mosavi (2010), Wu et al. (2010), Li et al. (2012), Shollo & Galliers (2013), Corte-Real et al. (2014), Hopken (2014), Rao (2015)).

This research illustrates that DM is an excellent tool for discovering intricate relationships that are often governed by the nonlinear functions among input and output variables (as well as solely among input variables). As such, using DM as an instrument to measure and evaluate the impact of KM on OR leads to many organizational benefits, some of which include the following:

- The ability of an organization to determine its 'OR Score';
- Practical ways of determining which KM activities lead to the largest gains in OR;
- Simulating the outcome on OR and inspecting the scenarios of certain KM initiatives;
- Identifying highly probable KM process-based reasons for the differences in performance between various levels of OR;
- Monitoring, by re-submitting new data, the performance of an organization with respect to OR; and
- Providing easier and more complex methods of analyzing OR than those offered by the tools based on classical statistics.
- The application of DM-based models can result in changes to organizational strategy that assure an organization that has achieved, or

is maintaining, a certain OR level making it well positioned not too fail in the future and perhaps to even take advantage of the market opportunities during the challenging and not-challenging for business times.

6.7.5 Chapter Summary

This chapter, based on the specific positioning of this work established in the introduction of this chapter, presented the findings of the research, focusing on answering the research questions and addressing the aim and objectives of this thesis. In addition to the presentation of findings and discussing their meaning, this chapter also dealt with contrasting this work with that of other authors.

6.8 Summary (Findings)

Chapter 6 presented the findings of this study that pertained to the DM-based component of this research. Specifically, that chapter was devoted to answering research questions #3, #4 and #5. The next chapter (Chapter 7) focuses on the conclusions to this research.

In Section 5.2, the business understanding aspect of this DM-based project was introduced and discussed. From the practitioner-based perspective, to which this research seeks to contribute, the following can be stated, in the business understanding context, as the summary of findings.

The DM tool proved to be an excellent tool for capturing the intricate and complex relationships between KM and OR.

The DM-based tool provided a way of determining which KM processes, and to what extent, made an organization resilient (based on the definition of OR used in this research); it also allows for the comparison of organizations with varying OR levels, making it possible to identify the KM processes that distinguish organizations when it comes to OR. There are many potential future applications of the tool and the findings generated by this research, including the use of the 'OR Score' to categorize organizations in order to anticipate their future performance.

Finally, the sophisticated and clear graphical user interface facilitated the use of the DM-based tool and made it easy to grasp the insights it offers.

Therefore, this research finds that DM-based tools have truly great potential for evaluating the impact of KM on OR.

CHAPTER SEVEN: CONCLUSIONS

7.1 Introduction

The aim of this research was to test the feasibility of using DM to assess the relationship between and impact of KM on OR.

This chapter explores the extent to which the aim of the research has been achieved in terms of answering the research questions. The contributions of this work to the contextual, methodological, empirical and organizational/professional areas are also examined.

The chapter closes with a discussion of some of the most apparent limitations of this research, suggestions for further research and concluding remarks.

7.2 Original Contributions of this Research

The original contributions that derive from the findings of this research relate to three main areas. First, this thesis makes methodological contributions relating to the development and application of the research methods; second, it makes contributions relating to understanding the concept of OR in relation to organizational performance and competitive advantage; and, finally, this research contributes to the professional/organizational field by the introduction of an OR model (originally presented in Section 3.5 and restated in the figure 7.2.1.1 below) that organizations may use to improve their resilience or to become more resilient organizations. A summary of the discussion that builds on the material presented in Section 3.5, on how the OR model may improve or lead to organizational resilience, follows in the next section.

7.2.1 Contextual Contributions

The research has made substantial contributions regarding the understanding of the impact of KM on OR, the relationships between KM and OR and methods that can be used to measure the impact of KM on OR. The understanding of the concept of OR has also been greatly improved as result of this work. Although not initially identified as a specific research question, one outcome of this research has been the development of a theoretical model of the resilient organization (this model was presented in Section 3.5 and is re-stated in the figure 7.2.1.1 below). The OR model (re-stated below) builds on the literature review conducted as part of this research and presents (what appears at the time of writing) a comprehensive and highly practical model (due to the considered constraints) for improvement or achievement of OR. The argument in support of the model is that the model implements many of the elements noted to contribute to OR. In addition, it also uses additional views (market, stakeholder and resource) as inputs to inform key environmental factors. While the elements supporting the OR models were discussed in Section 3.5, it is worth stating that the OR model seeks to improve or achieve OR through a systematic construction of key elements that include: data mining for environmental sensing and decision-making; KM components that seek to emphasize the KM as a part of the organization's strategy; four key inputs that expand the system to include other than KM-based components (components based on the following views: market, shareholder, resources and knowledge); OR enabling factors such as shared organizational principles and organizational culture; and the integrating processes facilitating learning and providing a feedback from DM to strategic KM. When implemented, the model addresses key factors of OR identified in the literature and expands the dynamic capabilities of an organization, a concept recently gaining on significance as a significant factor in organizational performance and, therefore in OR.

The contextual contributions made by this research were guided by the following research questions:

Research Question #1:

> What prior research exists regarding the application of DM with respect to KM and OR and the impact of KM on OR, and what are the known relationships between KM and OR?

Research Question #2:

> Can OR be measured pragmatically? Can the impact of KM on OR be measured pragmatically?

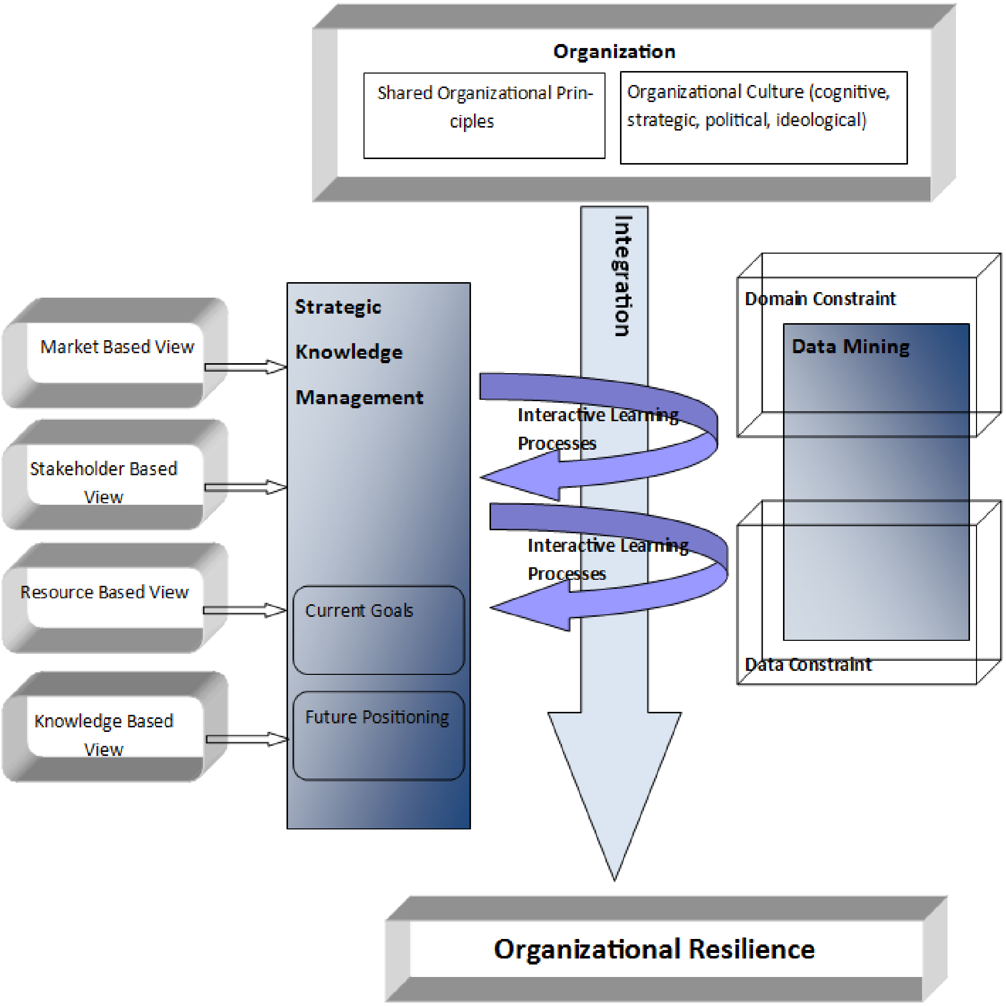

Organization
Shared Organizational Principles
Organizational Culture (cognitive, strategic, political, ideological)
Integration
Market Based View
Stakeholder Based View
Resource Based View
Knowledge Based View
Strategic Knowledge Management
Current Goals
Future Positioning
Interactive Learning Processes
Interactive Learning Processes
Domain Constraint
Data Mining
Data Constraint
Organizational Resilience

Fig. 7.2.1.1: The OR model

Prior to answering RQ#1, this research established, through the literature review (presented in Chapter 3) and a statistical analysis (presented in Appendix III), the similarities between organizational performance and OR. Those actions made it possible to equate organizational performance with OR for the purpose of conducting the literature review and this research as a whole. Establishing the correlation between the concept of organizational performance and OR means that this this research makes a contribution to the knowledge regarding OR.

As already mentioned in the literature review, only a very limited number of published works address RQ #1. Of the existing publications, the work of Choi et al. (2008) and Wu et al. (2010) use DM to evaluate the impact of KM on business performance. However, as opposed to this research, Choi et al. (2008)

investigated such impact in the context of KM strategy and not KM processes; furthermore, Wu et al. (2010) did not use OR as an evaluation metric. This research is, therefore, the first that considers evaluating the impact of KM processes on OR in a practical fashion, through the use of the KM-process-based frameworks of Burnett et al. (2004, 2013) and McKenzie and van Winkelen (2004). The use of these frameworks (presented in Sections 3.2.3, 3.2.4, 3.2.5 and 4.6) as a lens for the measurement of the impact of KM on OR is also innovative.

With regard to the approach of using the frameworks of Burnett et al. (2004, 2013) and McKenzie and van Winkelen (2004), no works were found that attempt to map KM processes onto the competence model with the intention of using the competence model to investigate the impact of KM on OR.

In relation to RQ #2, which is closely related to RQ #1 (as presented in Sections 3.3 and 3.4), there were numerous attempts to pragmatically measure OR and KM's impact on business, starting as early as the late 1990s (Horne (1997) and Horne & Orr (1998)). However, where this research differs is in the definition of OR: it links OR to organizational performance (OP), which builds on the research that addresses the measurement of OP as result of KM processes.

7.2.2 Methodological and Empirical Contributions

The primary focus of this research on methodological contributions is reflected in the following aim of this research:

> To test the feasibility of using DM to assess the relationship between and impact of KM on OR.

The aim of the research is supported by the following research questions:

Research Question #3:

> Which KM processes are the most influential for OR?

Research Question #4:

> Can a methodological approach be developed to examine the relationships between KM and OR, utilizing DM?

Research Question #5:

> Which are some of the main challenges when employing DM for the purpose of determining the impact of KM on OR?

Through the use of practical DM models based on the literature review, this research demonstrates how DM can be used methodically and empirically to measure the impact of KM on OR. The outcome of this research allows organizations to do the following:

- Arrive at a resilience score (called, in this research, the OR-Score) which is derived from the DM model, based on the replies to the questionnaire;
- Determine the KM processes that affect, either positively or negatively, OR and to what extent, in terms of probability, they do so (which, in fact, addresses the central purpose of this research);
- Compare the KM processes of resilient and non-resilient organizations to determine which KM activities are responsible for either low or high OR, and to what extent (in terms of probability);
- Inspect which KM processes are related to each other; and
- Determine the level of accuracy of the resultant DM model measuring the impact of KM on OR.

The outcome of this research is the first comprehensive and practical look at how DM can be used as a measurement instrument to measure the impact of KM on OR.

This research practically supports the claims of various other researchers (Brusilovski & Brusilovski (2008); Moyar & Gardner (2012); and Gullo (2015)) regarding DM's ability to generate useful contextual knowledge that is not easily obtained through the application of classical statistics. This research therefore adds to the body of knowledge, which is characterized by Corte-Real et al. (2014, pg. 176) as otherwise lacking, in that 'there is little understanding of how BI&A systems may effectively be used and create positive impact on the organization'. It is anticipated that the research described in this thesis will highlight the value of the DM as a tool for the data analysis to academic researchers.

In addition, this research utilized the methodological map presented in Chapter 6.7, which can be followed by organizations in order to improve their chances of success when carrying out DM-based projects. As was mentioned in Chapter 6.7, when discussing the findings related to RQ #5, there are a number of issues that can be encountered when conducting DM-based projects. The most common issues and challenges, including those encountered in this research, were presented with the hopes of smoothing out the DM process for future researchers and practitioners.

7.2.3 Organizational/Professional Contributions

The practical contributions of this work can be highlighted based on two main aspects: consideration of how DM can be used to assess the impact of KM processes on OR; and the introduction (in Section 3.5) of the OR model. With OR being a key to business success, it is imperative for professionals and organizations to be able to identify which factors make organizations resilient, as well as how to achieve OR.

In relation to the work included in this thesis, and given appropriate testing of the methods with suitable volumes of data, it could provide the basis for a guide which may be used to determine an OR Score (or level) based on the KM processes taking place within organizational contexts. At a higher granular level, the OR Score for a group of organizations or an industry may also be determined.

The model presented in Section 3.5 and restated in Section 7.2.1 contributes to such knowledge and understanding, as it provides a methodological way of improving or achieving OR through the expansion of the inputs to the model from KM process-based inputs only to include the consideration of market, stakeholder and resource-based views. Organizations may benefit from this work by becoming more resilient to the constantly changing business environment by reducing uncertainty and managing risk through the implementation of the OR model. Because of inclusion of various perspectives, the OR model, is expected to deliver meaningful and actionable results. Since the significant portion of the proposed OR model relies on DM, this work provides organizations with a structured DM implementation framework and highlights issues that organizations are likely to encounter in the use of DM models. With the help of this research, organizations now have a way of analyzing and improving their resilience.

In addition, with the help of this work, consulting companies or individual consultants may assist their end clients with both the implementation of the OR model, as well as implementation of the DM models and the analysis of the outcomes they produce.

7.3 Limitations

While this thesis successfully addressed its aims and objectives, there were many other aspects related to this project that were not addressed in the research.

Some of the shortcomings and/or concerns identified while conducting this research include the following:

Starr et al. (2003) discuss systematic resilience. This work, however, does not address the issue of the impact of organizational interdependencies on OR. It also does not consider the impact of KM on OR within a network. Rather, this work was focused on a single organization, without taking into account that organization's networked environment. As pointed out by Starr et al., such an environment can have a significant positive or negative impact on OR.

On a similar note, the work of McCann et al. (2009) investigated OR at multiple levels: individual, team, organization and industry, whereas this research only addressed the organizational level of resilience.

One of the limitations of this work is the lack of extensive focus on so-called KM-enablers. Typically, organizational structure, culture, leadership and IT-infrastructure are elements referred to as KM-enablers. While the importance of KM-enablers has been generally acknowledged in the successful implementation and management of KM initiatives, this work did not explicitly address KM-enablers. Rather, the view of this work is that KM-enablers have been 'factored in' into the effects and/or impacts of KM in an organization; that is, the effects of a knowledge-sharing culture and an extensive and up-to-date IT-infrastructure will positively impact the KM initiatives and, ultimately, the company's performance.

This work did not investigate the various strategies, tools and technologies that are used in KM. As stated by Haslinda and Sarinah (2009, pg. 189), when a resource-based view is employed (where knowledge viewed as an object), management should emphasize the building of a stock of knowledge and repositories to hold such knowledge. From the process-based view, then, the emphasis should be on knowledge creation, sharing and distribution.

Related to the insights generated by the DM algorithms, care must be taken with regard to the issues of respondent privacy, data security and the misuse of the information. While these issues were addressed in operational terms, they were not discussed at any great length in this research.

The use of a single vendor DM platform (SQL Server 2012 from the Microsoft Corporation) can be seen as a limiting factor when attempting to arrive at additional insights.

Finally, the limited volume of input data collected in response to the questionnaires represented a significant limitation to this research, which was addressed through a revision to the objectives established for the research. The need for such shift in the focus of the research was due to the fact that the amount of data collected could not support the construction of reliable models. Therefore, the focus of the research had to change from the pragmatic attempt to measure the OR to the theoretical approach illustrating the suitability of DM as a tool to measure the impact of KM-processes on OR.

7.4 Future Research

Although the research itself has achieved its aim of illustrating how DM can be successfully used to assess the impact of KM on OR, it also raises a further set of research questions that may be addressed by future research. The potential areas for future research have been identified primarily in relation to the methodological contributions made by this research.

Given the applied nature of this research, it is unsurprising that the suggested follow-up research is a logical extension of this study; several follow-up research projects can be suggested.

One of the most important suggestions would be the application of this study to the data set supporting such analysis. That is, the collection of a sufficient amount of data to obtain meaningful numerical results using the models presented in this work. Such a study, among other topics mentioned in this research, would allow for the identification of the key KM processes that make companies resilient; it could also allow for contrasting resilient and non-resilient organizations and identifying differences in KM processes.
In addition, given a sufficient amount of data, the DM algorithms considered could involve the association DM task. Such an investigation could lead to the identification of KM processes that are associated (that is, if one KM process occurs, then another process occurs as well); this could be particularly useful in identifying which KM attributes must occur together to make an organization resilient.

Beyond the concentration on the KM processes, the propositions above could involve the use of KM activities instead of KM processes; that is, each of the KM activities (corresponding to a single question in the questionnaire) could be analyzed using DM to determine its impact on the OR. Such work could be entirely based on the findings of this research, which focuses on groups of KM activities (competence areas) rather than single KM activity.

Finally, it would be worthwhile to apply the findings of this research to a sufficiently large dataset to learn which industries are the most resilient and which KM processes are responsible for resilience.

Clearly, applying the proposed research schemas to public companies would be beneficial as such an application would facilitate the validation of the findings of the DM algorithms based on the actual financial operational data that is available to the public.

From a practical standpoint, due to the nature of this thesis fulfilling the requirement for the professional doctorate degree, the author of this work intends to commercialize the OR model presented in this work (Section 3.5 and 7.2.1) within the next twelve months. To achieve this, the research presented in this thesis will need to be modified in a number of ways including the addition of questions related to newly added views, a reduction of KM-based questions, and the expansion of DM models to include the consideration of new views. In addition, the author intends to re-distribute a newly designed questionnaire (to a sample group to be determined) and undertake an organizational case study to illustrate and document the OR model's implementation. Because of these commercial plans, an embargo period of 36 months has been placed on this thesis.

7.5 Concluding Remarks

Using the DM method for the analysis of the impact of KM on OR is a very promising approach that should be given a great deal of attention by companies that wish to be resilient in order to adapt to changing business conditions. Consulting organizations should also consider the findings of this research as part of their consulting services, as doing so may provide additional value to their clients.

Due to the employment of DM in analyzing the impact of KM on OR, it is possible, as a result of this research, to consider such an impact in an innovative way. This innovative approach could include, among other things, measuring a company's OR level, determining the most and/or least effective OR and KM processes, identifying missing or underperforming KM processes and identifying the KM processes primarily responsible for an organization's resilience. Moreover, the impact of KM processes on the OR can be expressed with a numerical probability.

As stated by Davenport and Harris (2007, pg. 7), '[t]he questions that analytics can answer represent the higher-value and more proactive end of the spectrum' when looking at analytics to provide both a higher degree of intelligence and a higher level of competitive advantage.

As we move into increasingly uncertain times and socio-political environments, the need for organizations to understand their ability to become, and remain, resilient will be more important than ever. This research may well be among the first works in a burgeoning area of interest for both researchers and practitioners.

REFERENCES

Abbott, D. (2014). *Applied Predictive Analytics: Principles and Techniques for the Professional Data Analyst.* Indianapolis, IN: John Wiley & Sons, Inc.

Abdi, H., Valentin, D., & Edelman, B. (1999). *Neural Networks. Series: Quantitative Applications in the Social Sciences.* Sage Publications, Inc.

Abello, J., Pardalos, P. M., & Resende, M. (2002). *Handbook of massive data sets.* New York: Kluwer.

Adejuwon, A., & Mosavi, A. (2010). Domain Driven Data Mining-Application to Business. *IJCSI International Journal of Computer Science Issues, 7* (4), 2.

Aghdaie, M. H., Zolfani, S. H., & Zavadskas, E. K. (2014). Synergies of data mining and multiple attribute decision making. *Contemporary Issues in Business, Management and Education. Procedia – Social and Behavioral Sciences, 110*, 767-776.

Akkaya, G. C, & Uzar, C. (2011). Data Mining in Financial Application. *Journal of Modern Accounting and Auditing, 7* (12), 1362-1367.

Alavi, M., & Leidner, D.E. (2001). Review: Knowledge Management And Knowledge Management Systems: Conceptual Foundations and Research Issues. *MIS Quarterly, 25* (1), 107-136.

Alsultanny, Y. (2011). Selecting a suitable method of data mining for successful forecasting. *Journal of Targeting Measurement and Analysis for Marketing, 19*, (¾), 207-225.

Ananthakrishna, R., Chaudhuri, S., & Ganti, V. (2002). Eliminating Fuzzy Duplicates in Data Warehouses. *Procieedings of the 28th VLDB Conference*, Hong Kong, China.

Anderson, D. R., Sweeney, D. J., Williams, T. A. (2003). *Essentials of Statistics for Business and Economics* (3rd Edition). Thomson, South-Western.

Andone, I. I. (2009). Measuring the Performance of Corporate Knowledge Management Systems. *Informatica Economica, 13* (4), 25-31.

Andonie, R. (2010). Extreme Data Mining: Inference from Small Datasets. International *Journal. of Computer Communications & Control, 5* (3), 280-291.

Andonie, R., Fabry-Asztalos, L., Magill, L., & Abdul-Wahid, S. (Eds.). (2007). A new fuzzy ARTMAP approach for predicting biological activity of potential HIV-1 protease inhibitors. *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM).* San Jose, CA: I.C.S. Press.

Ankerst, M. (2002). Panel the Perfect Data Mining Tool: Interactive or Automated? *ACM SIGKDD Exploration Newsletter, 4* (2), 110-111.

Anon. (2006). Knowledge Management: your company's completive advantage. *Finweek. Sandton,* 1-64.

Anon. (2006). The difficulties of measuring KM. *Knowledge Management Review, 9*, 6-8.

Anon. (n.d). Advantages and disadvantages of different types of observational studies. University of North Carolina. Retrieved from Lecture Notes Online Web site: http://ssw.unc.edu/mch/node/221. [Retrieved on Oct. 30, 2016.]

Anon. (n.d.). Questionnaire design. University Library, Loughborough University http://www.lboro.ac.uk/media/wwwlboroacuk/content/library/downloads/advicesheets/questionnaire%20no%20logo.pdf . [Retrieved on Nov. 14, 2011.]

Armistead, C. (1999). Knowledge management and process performance*.* *Journal of Knowledge Management, 3* (2), 143-154.

Arnott, D. & Pervan, G. (2008). Eight key issues for the decision support systems discipline. *Decision Support Systems, 44*, 657-672.

Baicoianu, A., & Dumitrescu, S. (2010). Data Mining Meets Economic Analysis: Opportunities and Challenges. Bulletin of the Transilvania University of Brasov. Series V. *Economic Sciences, 3* (52).

Banu, A.B., Balamurugan, S.A., & Thirumalaikolundusubramanian, P. (2014). Detection of dechallenge in spontaneous reporting systems: A comparison of Bayes methods. *Indian J Pharmacol, 46*, 277-80.

Barclay, R. O., & Murray, P. C. (1997). *What is Knowledge Management?* A Knowledge Praxis. Retrieved from http://www.providersedge.com/docs/km_articles/What_Is_Knowledge_Management.pdf . [Retrieved on June 3, 2010.]

Barney, J.B. (1992). Integrating organizational behavior and strategy formulation research: a resource based analysis. In P. Shrivastava, A. Huff, & J. Dutton (Eds.), *Advances in Strategic Management, 8*, 39-62. Greenwich, CT: JAI Press.

Barney, J. B. (1995). Looking inside for competitive advantage. *Academy of Management Executive, 9* (4).

Benaroch, M. (2002). Managing Information Technology Investment Risk: A Real Options Perspective. *Journal of Management Information Systems,* 19 (2), 43-84.

Benaroch, M., Lichtenstein, Y., & Robinson, K. (2006). Real Options in Information Technology Risk Management: An Empirical Validation of Risk-Option Relationships*. MIS Quarterly, 30* (4), 827-864.

Benbya, B., Passiante, G., & Belbaly, N. A. (2004). Corporate portal: a tool for knowledge management synchronization. *International Journal of Information Management, 24* (3), 201-220.

Benn, P. (2011). *Managing for Resilience*. Business Continuity Institute.

Berson, A., Smith, S., & Thearling, K. (2000). *Building data mining applications for CRM.* New York: McGraw-Hill.

Berthold, M., & Hand, D. J. (1999). *Intelligent Data Analysis.* Springer-Verlag Berlin Heidelberg.

Bierly, P., & Chakrabarti, A. (1996). Generic knowledge strategies in the US pharmaceutical industry. *Strategic Management Journal, 17* (4), 123-135.

Birkinshaw, J., Gibson, C. (2004). Building Ambidexterity Into An Organization. *MIT Sloan Management Review. 45*(4), 47-55.

Bissantz, N., & Hagedorn, J. (2009). Data Mining. *Business & Information Systems Engineering,* 1.

Bocharov, A., & Lind, J. (2005). Data Mining Reloaded. *SQL Server Magazine*.

Borgatti P. S. & Carboni I. (2007) On Measuring Individual Knowledge in Organizations. *Organizational Research Methods, 10*, 449-462.

Bos, S., & Chug, E. (1996). Using weight decay to optimize generalization ability of a perception. *Proceedings of the 1996 International Conference on Neural Networks, IEEE*, 241-246.

Bose, R., & Sugumaran, V. (2003). Application of knowledge management technology in customer relationship management. *Knowledge and Process Management, 10* (1), 3-17.

Braes, B., & Brooks, D. (2010). Organizational Resilience: A Propositional Study to Understand and Identify the Essential Concepts. *Proceedings of the 3rd Australian Security and Intelligence Conference*. Perth Western Australia.

Brown, B., Chui, M., & Manyika, J. (2011). Are you ready for the era of 'big data'? McKinsey Quarterly. http://www.mckinsey.com/insights/strategyare_you_ready_for_the_era_of_big_data. [Retrieved on Jun. 14, 2015].

Brusilovsky, D., & Brusilovski, E. (2008). Data Mining: The Means to a Competitive Advantage. Business Intelligence Solutions. Retrieved from: http://www.bisolutions.us/The-Means-to-a-Competitive-Advantage.php. [Retrieved on Sep. 1, 2015].

Brynjolfsson, E., Hitt, L. M. & Kim, H. H. (2011). Strength in Numbers: How Does Data-Driven Decision-making Affect Firm Performance? Retrieved from http://ssrn.com/abstract=1819486 [Retrieved on Nov. 2, 2013].

Burnett, S., Illingworth, L., & Webster, L., 2004. Knowledge auditing and mapping: a pragmatic approach. *Knowledge and Process Management, 11* (1), 25-37.

Burnett, S., Williams, D., & Grinnall, A. (2013). The strategic role of knowledge auditing and mapping: An organisational case*. Knowledge and Process Management*, 20 (3), 161-176.

Burnett, S., Williams, D., & Illingworth, L. (2013). Reconsidering the Knowledge Audit Process: Methodological Revisions in Practice. *Knowledge and Process Management*, *20* (3), 141-153.

Cai, W., Yang, C., Smarandache, F., Vladareanu, L., Li, Q., Zou, G. & Li, X. (2013). *Extenics and Innovation Methods*. Boca Raton, FL: CRC Press.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and Quasi-Experimental Designs for Research*. Rand McNally: Skokie, IL.

Cao, L., & Zhang, C. (2006). Domain-Driven Data Mining: A Practical Methodology. *International Journal of Data Warehousing & Mining, 2* (4), 49-65.

Carlucci, D., & Schuima, G. (2006). Knowledge Asset Value Spiral: Linking Knowledge Assets to Company's Performance. *Knowledge and Process Management, 13* (1), 35-46.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability And Validity Assessment. Series: Quantitative Applications in the Social Sciences*. Inc. Thousand Oaks, CA: Sage Publications, Inc.

Carrier, C. G., & Povel, O. (2003). Characterizing data mining software. *Intelligent Data Analysis,* 7, 181-192.

Chae, B., Yang, C., Olson, D., & Shew, C. (2014). The impact of advanced analytics and data accuracy on operational performance: A contingent resource based theory (RBT) perspective. *Decision Support Systems, 59*, 119 – 126.

Chantal & Chantal. *(2015). Data Mining and Predictive Analytics*. Wiley Series on Methods and Applications in Data Mining. 2nd Edition. Hoboken, New Jersey: John Wiley & Sons, Inc.

Chcillar, S. K., & Khehra, B. S. (2012). Decision Tree approach to predict the Gender wise response on the volume of applications in government organizations for recruitment process. *Int. J. Computer Technology & Applications, 3* (6), 2018-2021.

Chemchem, A., & Drias, H. (2015). From data mining to knowledge mining: Application to intelligent agents. *Expert System with Applications, 42*, 1436-1445.

Chen, M. J. (1996). Competitor Analysis and Interfirm Rivalry: Toward a Theoretical Integration. *Academy of Management Review, 21* (1), 27-36.

Chen, M. Y., Huang, M. J., & Cheng, Y. C. (2009). Measuring knowledge management performance using a competitive perspective: An empirical study. *Expert Systems with Applications, 36*, 8449-8459.

Chen, X., & Siau, K. K. (2012). Effect of Business Intelligence and IT Infrastructure Flexibility on Organizational Agility. *Thirty Third International Conference on Information Systems*, Orlando.

Cheng, H., Lu, Y., & Sheu, C. (2009). An ontology-based business intelligence application in a financial knowledge management system. *Expert Systems with Applications, 36*, 3614 – 3622.

Choi, B., & Lee, H. (2003). An empirical investigation of KM styles and their effect on corporate performance. *Information and Management, 40* (5), 403-417.

Choi, B., Poon, S. L., & Davis, J. G. (2008). Effects of knowledge management strategy on organizational performance: A complementarity theory-based approach. *The International Journal of Management Science Omega, 36*, 235-251.

Chou, C. (2011). A Framework for Aligning Strategic Positioning and Knowledge Management System. *Information Technology Journal, 10* (8), 1594-1600.

CIA – The World Fact Book: United States. Retrieved from https://www.cia.gov/library/publications/the-world-factbook/geos/us.html [Retrieved on October 23, 2013].

Clark, L. (2013). No questions asked: Big data firm maps solutions without human input. Wired, 16.

CNBC: www.cnbc.com/id/100471829 [Retrieved on Nov. 6, 2013].

CNN: money.cnn.com/magazines/fortune/best-companies/2012/midsized.html [Retrieved on Nov. 6, 2013].

Cockram, D., & Heuvel. C. (2012). BCI Partnership. Retrieved from https://issuu.com/steelhenge/docs/bci_organisationalresiliencepaper [Retrieved on Apr. 5, 2013.]

Converse, J. M., & Presser, S. (1986). *Survey Questions Handcrafting The Standardized Questionnaire*. Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 63-001. Beverly Hills: Sage Publications, Inc.

Corte-Real, N., Ruivo, P., & Oliveira, T. (2014). The diffusion of business intelligence & analytics (BI&A): A systematic mapping study. *Procedia Technology*, *16*, 172-179.

Coutu, D. L. (2002). How Resilience Works. *Harvard Business Review At Large,* 46.

Cox, S. M., & Harper, M. (2012). Target: The Challenge of Data Mining. *Journal of Critical Incidents, 6*.

Creswell, J. W. (2003). *Research Design*. Qualitative, Quantitative, and Mixed Methods Approaches. 2nd Edition. Thousand Oaks: Sage Publications, Inc.

Creswell, J. W. (2009). *Research Design*. Third Ed. Thousand Oaks, CA: Sage Publications, Inc.

Cronbach, L. J. (1971). Test Validation. In R. L. Thorndike (ed.) *Educational Measurement* (443-507). Washington DC: American Council on Education.

Crook, R. T., Combs, J. G., Todd, S. Y., Woehr, D. J., & Ketcher Jr. D.J. (2011). Does Human Capital Matter? A Meta-Analysis of the Relationship Between Human Capital and Firm Performance. *Journal of Applied Psychology,* 96 (3), 443-456.

Davenport, T. (2010). Business Intelligence and Organizational Decision. International *Journal of Business Intelligence Research (IJBIR),* 1, 1-12.

Davenport, T. H., & Harris, J. G. (2007). *Competing on Analytics.* Boston, MA: Harvard Business School Press.

Davenport, T. H., Harris, J. G., & Morison, R. (2010). *Analytics at Work.* Boston, MA: Harvard Business School Press.

Davenport, T., & Prusak, L. (1998). *Working knowledge: How Organizations Manage What They Know.* Boston, MA: Harvard Business School Press.

De Ville, B. (2001). *Microsoft Data Mining: Integrated Business Intelligence for e-Commerce and Knowledge Management.* Butterworth-Heinemann, Woburn, MA, USA.

Deloitte. Mid-market perspectives. (2012) report on America's economic engine.

Dennis, S., (n.d.). How to design a questionnaire / survey. The University of New South Wales, Sidney, Australia. Retrieved from http://www.powershow.com/view/3ae445-YWQ5Y/How_to_design_a_questionnaire_survey_powerpoint_ppt_presentation [Retrieved on Mar. 17, 2013].

Desouza, K. C. (2006). The difficulties of measuring KM. Q&A with Dr. Kevin C. Desouza, The Information School, University of Washington. *Knowledge Management Review*, *9* (5), 6-7.

DiBella, A. and Nevis, E. 1998. *How Organizations Learn: An Integrated Strategy for Building Learning Capability.* San Francisco, CA: Jossey-Bass Press.

Dolfsama, W. & Leydesdorff, L, (2008). "Medium-tech" industries may be of greater importance to a local economy than "High-tech" firms: New methods for measuring the knowledge base of an economic system. *Medical Hypotheses, 71* (3), 330-334.

Donaldson, T., & Preston, L. E. (1995). The Stakeholder Theory of the Corporation: Concepts, Evidence, and Implications. *The Academy of Management Review, 20* (1), 65-91.

Evans, J. R., & Lidner, C. H. (2012). Business Analytics: The Next Frontier for Decision Science. Decision Science Institute. Retrieved from http://www.cbpp.uaa.alaska.edu/afef/business_analytics.htm [Retrieved on January 29, 2014].

Fichman, R. G. (2004). Real Options and IT Platform Adoption: Implications for Theory and Practice. *Information Systems Research, 15* (2), 132-154.

Field, A. (2006). Reliability Analysis. C8057 (Research Methods II): Reliability Analysis. Retrieved from 'http://www.statisticshell.com/docs/reliability.pdf [Retrieved on Apr. 4, 2013].

Fink, K., & Ploder, K. (2007). A Comparative Study of Knowledge Processes and Methods in Austrian and Swiss SMEs. *ECIS 2007 Proceedings*, Paper 193.

Folorunso O., Ogunde A. O. (2005). Data mining as a technique for knowledge management in business process redesign. *Information Management & computer Security*, *13* (4), 274 – 280.

Fowler, J. & Floyd Jr. (2014). *Survey Research Methods*. London, UK: Sage Publications, Inc.

Frappaolo, F. (2006). *Knowledge Management*. Capstone Publishing Ltd. 1998.

Frary, R.B. (2002). A Brief Guide to Questionnaire Development. Indiana University.
http://www.indiana.edu/~educy520/sec5982/week_3/questionnaire_development_frary.pdf [Retrieved on Jun. 12, 2013].

Freeman, R.E. (2010). *Strategic Management*: A stakeholder Approach. New York, NY: Cambridge University Press

Freeman, R. E., & McVea, J. (2001). *A stakeholder approach to strategic management.* SSRN Electronic Journal.

Fricke, M. (2007). The knowledge Pyramid: A Critique of the DIKW Hierarchy. *Journal of Information Science, 35* (2), 131-142.

Frid, R. (2003). A Common KM Framework For The Government Of Canada: Frid Framework For Enterprise Knowledge.

Friedman, M. (2005). *Organisational Resilience*. Accountancy SA, 24.

Fuchs, M., Hopken, W., & Lexhagen, M. (2014). Big data Analytics for knowledge generation in tourism destinations – A case from Sweden. *Journal of Destination Marketing & Management, 3*, 198-209.

Gabriel, Y. (2000). *Storytelling in Organisations*. Oxford: Oxford University Press.

Galbraith, J.K. (1969). *The New Industrial State*. Harmondsworth: Penguin.

Gamez, J., Modave, F., & Kosheleva, O. (2008). Selecting the most representative sample is NP-hard: Need for expert (fuzzy) knowledge. *IEEE International Conference on World Congress on Computational Intelligence*, 1069-1074.

Gartner Research. Retrieved from http://www.gartner.com/it-glossary/smbs-small-and-midsize-businesses [Retrieved on Oct. 31, 2013.]

Gehl, R. W. (2015). Sharing, knowledge management and big data: A partial genealogy of the data scientist. *European Journal of Cultural Studies, 18* (4-5), 413-428.

Gliem, J. A., & Gliem, R. R. (2003). Calculating, Interpreting, and Reporting Cronbach's Alpha Reliability Coefficient for Likert-Type Scales. Midwest Research to Practice Conference in Adult, Continuing, and Community Education.

Grant, R. M. (1996). Prospering in dynamically-competitive environments: Organizational Capability as knowledge integration. *Organization Science, 7*, 375-387.

Green, A. (2004). Prioritization of sources of intangible assets for use in enterprise balance scorecard valuation models of information technology (IT) firms (unpublished doctoral dissertation). George Washington University, Washington, DC.

Green, A. (2006). Knowledge Valuation. The starting block: enterprise (business) intelligence – evolving towards knowledge valuation. *The journal of information and knowledge management systems, 36* (3), 267-277.

Grus, J. (2015). Data Science from Scratch. O'Reilly. Sebastopol, CA.

Gullo. F., (2015). *From Patterns in Data to Knowledge Discovery: What Data Mining Can Do.* Physics Procedia *62*, 18-22.

Gupta, A., & McDaniel, J. (2002). Creating Competitive Advantage By Effectively Managing Knowledge: A Framework for Knowledge Management. *Journal of Knowledge Management Practice*.

Halavi, A. L., McCarthy R. V., Aronson, J. E. (2006). Knowledge management and the competitive strategy of the firm. *The Learning Organization, 13* (4), 384-397.

Hamel, G., & Valikangas, L. (2003). The Quest for Resilience. *Harvard Business Review*.

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques*. 3rd Ed. Waltham, MA, USA.

Hand, D. J. (2007). Principles of Data Mining. *Drug Safety, 30* (7), 621-622

Handzic, M. (2009). Evaluating KMS Effectiveness for Decision Support. Preliminary Results. In W.R. King (ed.), *Knowledge Management and Organizational Learning, Annals of Information Systems* (4), Springer Science+Business Media.

Harlow, H. (2008). The effect of tacit knowledge on firm performance. *Journal of Knowledge Management*, *12 (1),* 148-163.

Hartwig, F., & Dearing, B. E. (1979). *Exploratory Data Analysis. Series: Quantitative Applications in the Social Sciences*. London: Sage Publications, Inc.

Harvard University. *Research Methods*. Retrieved from: isites.harvard.edu/fs/docs/icb.topic851950.files/Research%20Methods_Some%20 Notes.pdf [Retrieved on Aug. 14, 2016].

Haslinda, A., & Sarinah, A. (2009). A Review of Knowledge Management Models. *The Journal of International Social Research, 2* (9).

Heinrichs, J. H., & Lim, J. S. (2003). Integrating web-based data mining tools with business models for knowledge management. *Decision Support Systems, 35,* 103-112.

Hershel, R. & Yermish, I. (2009*).* Knowledge Management in Buiness Intelligence. In W.R. King (ed.) *Knowledge Management and Organizational Learning, Annals of Information Systems* 4, Springer Science+Business Media.

Hess, D. R. (2004). How to Write an Effective Discussion*. Respiratory Care, 49* (10), 1238-1241.

Holmstrom, L., & Koistinen, P. (1992). Using additive noise in backpropagation training. *IEEE Transactions on Neural Networks,* 3, 24-38.

Hopken, W., Fuchs, M., Keil, D., & Lexhagen, M. (2011). The knowledge destination – A customer information-based destination management information system. In R. Law, M. Fuchs, & F. Ricci (Eds.), *Information and communication technologies in tourism* (pp. 417 – 429), New York: Springer.

Hopkins, B., & Schadler, T. (2015). Digital insights as the new currency of business. *KM World, 22*, 10 – 11.

Horne III, J. F. (1997). The coming age of organizational resilience. *Business Forum, 22* (2-3), 24

Horne III, J. F., & Orr, J. E. (1998). Assessing Behaviors That Create Resilient Organizations. *Employment Relations Today, 24* (4), 29-39.

Huang, Y. F., Wu, W. W., & Lee, Y. T. (2008). Simplifying essential competencies for Taiwan civil servants by using the rough set approach. *Journal of the Operational Research Society, 59* (2), 259-266.

Hughes, L. P., & Holbrook, J. A. D. (1998). Measuring Knowledge Management: A New Indicator of Innovation In Enterprises. *CPROST Report* 98-02.

Hussain, F., Lucas, C., & Asif, A.M. (2004) Managing Knowledge Effectively. *Journal of Knowledge Management Practice.*

Husted, S. & Michailova, S. (2002). Diagnosing and fighting knowledge-sharing hostility. *Organizational Dynamics, 31*, 60-73.

Iacobucci, D., & Duhachek, A. (2003). Advancing Alphas: Measuring Reliability With Confidence. *Journal of Consumer Psychology, 13* (4), 478-487.

Ibrahim, F. & Reid, V. (2009). What is the Value of Knowledge Management Practices? *Electronic Journal of Knowledge Management, 7* (5), 567-574.

Intelligent Risk Systems (iJet). (2008). Business Resiliency For The Global Marketplace: Transforming Operating Risk Into Competitive Advantage. Retrieved from https://www.ijet.com/sites/default/files/Business_Resiliency_for_the_Global_Marketplace.pdf [Retrieved on Nov. 13, 2014.]

Isik, O., Jones, M. C., & Sidorowa, A. (2013). Business intelligence success: The roles of BI capabilities and decision environments. *Information & Management, 50*, 13-23.

Janus, P., & Misner, S. (2011). *Building Integrated Business Intelligence Solutions with SQL Server 2008 R2 & Office 2010.* The McGraw-Hill Companies.

Jackson, J. (2002). Data Mining: A Conceptual Overview. *Communications of the Association for Information Systems,* 8, 267-296.

Johnson, T., & Owens, Linda. (2003). Survey Response Rate Reporting In The Professional Literature. American Association for Public Opinion Research - Section on Survey Research Methods. http://www.amstat.org/sections/SRMS/Proceedings/y2003/Files/JSM2003-000638.pdf. [Retrieved on Sep. 1, 2016.]

Kalton, G. (1983). *Introduction To Survey Sampling*. Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 35-001. Beverly Hills: Sage Publications, Inc.

Kamara, J. M., Anumba, C. J., & Carrillo, P. M. (2002). A CLEVER approach to selecting a knowledge management strategy. *International Journal of Project Management,* 20 (3), 205-211.

Kankanhalli, A., & Tan, B. C. Y. (2004). A Review of Metrics for Knowledge Management Systems and Knowledge Managements Initiatives. *Proceedings of the 37th Hawaii International Conference of System Sciences, 8*.

Kaplowitz, M. D., Hadlock, T. D., & Ralph, L. (2004) A Comparison Of Web And Mail Survey Response Rates. *Public Opinion Quarterly, 68* (1), 94–101.

Karim, M., & Rahman, R. M. (2013). Decision Tree and Naïve Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing. *Journal of Software Engineering and Applications, 6*, 196-206.

Karystinos, G. N., & Pados, D. A. (2000). On overfitting, generalization, and randomly expanded training sets*. IEEE Transactions on Neural Networks, 5*, 1050-1057.

Kipley, H. D., Lewis, O. A., & Hlem R. (2008). Achieving Strategic Advantage and Organizational Legitimacy for Small and Medium Sized NFPs Through the *Implementation of Knowledge Management. Renaissance Quarterly*, 3 (3), 21, 22.

King, W. R. (2009). *Knowledge Management and Organizational Learning.* Annals of Information Systems, 4. NY: New York. Springer Science+Business Media.

Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography, 3* (3), 262-267.

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shift. *Big Data & Society,* 1-12.

Kogut, B., & Zander, U. (1992). Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology. *Organization Science, 3* (3), 383-397.

Koopman, C. (2013). *Genealogy as Critique: Foucault and the Problems of Modernity*. Bloomington, IN: Indiana University Press.

Kopelko, M., Jimenez, D. P., & Cirado, J. R. (2009). *Intangible Assets and Efficiency* (Doctoral Dissertation). University Autonoma de Barcelona.

Kowalczyk, M., Buxmann, P., & Besier, J. (2013). Investigating Business Intelligence And Analytics From A Decision Process Perspective: A Structured Literature Review. *Proceedings of the 21st European Conference on Information Systems*.

Krauss, S., & Eric. (2005). Research Paradigms and Meaning Making: A Primer. *The Qualitative Report*, 10, 758-770.

Kuhn Max & Johnson Kjell. (2016). *Applied Predictive Modeling*. Springer New York Heidelberg Dordrecht London.

Kulkarni, U., & Freeze, R. (2010). Measuring knowledge management capabilities. In *Encyclopedia of Knowledge Management, 1*, 1090-1100.

Kumari, M. (2011). Data Driven Data Mining to Domain Driven Data Mining. Global *Journal of Computer Science and Technology, 11* (23).

Lado, A., & Wilson, M. (1994). Human resource systems and sustained competitive advantage: a competency-based perspective. *Academy of Management Review, 19*, 699-727.

Lamont, J. (2015a). Creating a cohesive customer experience. *KM World*, 8-9.

Lamont, J. (2015b). Text analytics broadens its reach. *KM World*, 24 (7).

Larose, D. T. (2005). *Discovering knowledge in data*. New Jersey: John Wiley & Sons.

Larose, D.T., & Larose, D. C. (2015). *Data Mining and Predictive Analytics.* New Jersey: John Wiley & Sons

Larose, D.T., & Larose, D. C. *(2015). Data Mining and Predictive Analytics*. Wiley Series on Methods and Applications in Data Mining. 2nd Edition. Hoboken, New Jersey: John Wiley & Sons, Inc.

Larson, B. (2009). *Delivering Business Intelligence with Microsoft SQL Server 2008*. McGraw-Hill Companies.

Larson, B. (2012). *Delivering Business Intelligence with Microsoft SQL Server 2012* (Third Edition). McGraw-Hill Companies.

Lau, H. C. W., Choy, W. L., Law, P. K. H., Tsui, W. T. T., & Choi, L. C. (2004). An intelligent Logistics Support System for Enhancing the Airfreight Forwarding Business. *Expert Systems, 21* (5).

Lauck, J. K. (2013). Why the Midwest Matters*. The Midwest Quarterly*, *108* (112), 165-185 .

Law, C. C. H., & Ngai, E. W. T. (2003). An empirical study of the effects of knowledge sharing and learning behaviors on firm performance. *Expert Systems with Applications, 25* (2), 155-164.

LeBlanc, P., Moss, J. M., Sarka, D., Ryan, D. (2015). *Applied Microsoft Business Intelligence.* Wiley.

Lee, K.C., Lee, S., & Kang, I. W. (2005). KMPI: measuring knowledge management performance*, Information and Management, 42* (3), 469-482.

Lee M-C. (2008). Linkage Knowledge process and Business Process: A case study in China Motor Corporation. *International Conference on Convergence and Hybrid. Information Technology*, 407-412.

Legnick-Hall, C. A., & Beck, T. E. (2005). Adaptive Fit Versus Robust Transformation: How Organizations Respond To Environmental Change. *Journal of Management, 31*, 738.

Leung, C. K., & Joseph, K. W. (2014). Sports data mining: predicting results for the college football games. *Procedia Computer Science, 35*, 710-719.

Levin, A. K. (2006). Study design III: Cross-sectional studies. *Evidence-Based Dentistry*, 7, 24-25.

Li, X., Zhu, Z., & Pan, X. (2010). Knowledge Cultivating for Intelligent Decision Making in Small & Middle Businesses. *Procedia Computer Science, 1* (1), 2479-2488.

Liebowitz, J. (Ed.) (1999). *Knowledge Management Handbook*. Boca Raton: CRC Press.

Liebowitz, J., & Suen, C. (2000). Developing knowledge management metrics for measuring intellectual capital. *Journal of intellectual capital, 1* (1), 54-67.

Lina, C., & Tsen, S. M. (2005). Bridging the implementation gaps in the knowledge management system for enhancing corporate performance. *Expert Systems with Applications, 29* (1), 163-173.

Liu, Y., Starzyk, J. A., & Zhu, Z. (2008). Optimized approximation algorithm in neural networks without overfitting. *IEEE Transactions on Neural Networks, 19* (6), 983-995.

Lubatkin M. H., Simsek, Z., Ling, Y., Veiga J. F. (2006). Ambidexterity and Performance in Small-to Medium Sized Firms: The Pivotal Role of Top Management Team Behavioral Integration. *Journal of Management, 32* (5), 646-672.

Luo, J., Fan, M., & Zhang, H. (2012). Information technology and organizational capabilities: A longitudinal study of the apparel industry. *Decision Support Systems, 53*, 186 – 194.

MacLennan, J., Tang, Z., & Crivat, B. (2009). *Data Mining with Microsoft SQL Server 2008*. Indianapolis, IN: Wiley Publishing, Inc.

Mahdaviani, K., Mazyar, H., Majidi, S., & Saraee, H. (2008). A method to resolve the overfitting problem in recurrent neural networks for prediction on complex system's' behavior. In IJCNN'08: *Proceedings of the 2008 International Joint Conference on Neural Networks,* 3723-3728.

Malhorta, Y. (1998). Knowledge Management, Knowledge Organizations & Knowledge Workers: A View from the Front Lines. *Maeil Business Newspaper*.

Mallak, L. (1998). Putting organizational resilience to work. *Industrial Management, 40* (6).

Mandrai, Priyanka & Barskar, Raju. (2013). A Survey of Conceptual Data Mining and Applications. *International Journal of Computer Science and Information Security.* 11(5), 17 – 23.

Margo, H. (2004). Data-mining algorithms in Oracle9i and Microsoft SQL Server. Insights from industry Emerald series. *Campus-Wide Information Systems, 21* (3), 132-138.

Marr, B., Gupta, O., Pike, S., & Roos, G. (2003). Intellectual capital and knowledge management effectiveness. *Management Decisions, 41* (8), 771-781.

McAdam & McCreedy. (1999). A critical review of Knowledge Management models. The *Learning Organization*, 6 (3)

McCann, J., Selsky, J., & Lee, J. (2009). Building Agility, Resilience and Performance in Turbulent Environments. *People & Strategy*, 32, 3.

McCusker, K. & Gunaydin, S. (2015). Research using qualitative, quantitative or mixed methods and choice based on the research. *Perfusion, 30* (7) 537-542.

McDargh, E. (2003) Mastering resilience skills for off-the-chart results. *Management Quarterly*, 44, 1.

McElroy, M. W. (2003). *The new knowledge management*. KMCI Press/ Butterworth Heinemann.

McKenzie, J., & Winkelen, C. (2004). *Understanding the Knowledgeable Organization*. London: Thomson.

Microsoft Corporation. 'Cross Validation (Analysis Services – data Mining).' Retrieved from: https://msdn.microsoft.com/en-us/library/bb895174(d=printer,v=sql.110).aspx [Retrieved on Aug. 3, 2016.]

Microsoft Corporation. 'Cross-Validation (SQL Server Data Mining Add-ins).' Retrieved from: https://msdn.microsoft.com/eu-us/library/dn282375(d=printer).aspx [Retrieved on Aug. 4, 2016.]

Microsoft Corporation. 'Lift Chart (Analysis Services – Data Mining)' Retrieved from: https://msdn.microsoft.com/en-us/library/ms175428(v=sql.110).aspx [Retrieved on Jun. 15, 2016.]

Microsoft Corporation. 'Microsoft Clustering Algorithm' Retrieved from: https://msdn.microsoft.com/en-us/library/ms174879(v=sql.110).aspx [Retrieved on Jun. 29, 2014.]

Microsoft Corporation. 'Microsoft Neural Algorithm Technical Reference.' Retrieved from:

https://msdn.microsoft.com/en-us/library/ms174941(v=sql.110).aspx [Retrieved on Jul. 24, 2016.]

Microsoft Corporation. https://msdn.microsoft.com/en-us/library/cc645901(v=sql.110).aspx [Retrieved on Jul. 24, 2016.]

Ming-Chang, M. (2008). Linkage Knowledge process and Business Process: A case study in China Motor Corporation, 2008 *International Conference on Convergence and Hybrid Information Technolog*y, 407-412.

Moayer, S., & Gardner, S. (2012). Integration of data mining with a Strategic Knowledge Management framework: A platform for competitive advantage in the Australian mining sector. (IJACSA) *International Journal of Advanced Computer Science and Applications, 3* (8), 67 – 72.

Morales, R.D., & Wang, J. (2010). Forecasting cancellation rates for services booking revenue management using data mining. *European Journal of Operation Research*, *202*, 554-562.

Murray, P. (2002). Knowledge management as a sustained competitive advantage. *Ivey Business Journal, 66* (4), 6-71.

Narmad, V. (2014). Comparison of Association Rules, Clustering and Decision Tree Data Mining Models' Accuracy: A Case Study of Birth Registration E-governance data. *Indian Journal of Applied Research, 4* (9).

Natek, S., & Zwilling, M. (2013). Data Mining For Small Student Data Set – Knowledge Management System For Higher Education Teachers. *Management, Knowledge and Learning – International Conference*. Zadar, Croatia.

Natek, S., & Zwilling, M. (2014). Student data mining solution-knowledge management system related to higher education institutions. *Expert Systems with Applications*, 41, 6400-6407.

Ngai, E. W. T., Xiu, Li., & Chau, D. C. K. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, 36, 2592-2602.

Nickols, F.W. (2012). The knowledge in knowledge management. http://www.nickols.us/Knowledge_in_KM.htm. [Retrieved on Feb. 20, 2015.]

Nonaka, I. & Takeuchi, H. (1995). *The Knowledge-Creating Company*. Oxford: Oxford

Nonaka, I. (1991). The knowledge creating company. *Harvard Business Review, 69*, 96-104.

Oliveira, M. & Goldoni, V. (2006). Metrics for knowledge management process. *AMCIS Proceedings*, Paper 217.

Olszak, C. M., & Ziemba, E. (2007). Approach to Building and Implementing Business Intelligence Systems. *Interdisciplinary Journal of Information, Knowledge and Management*, 2, 135 - 148.

Patton, R. J. (2007) Metrics for Knowledge-Based Project Organizations. S.A.M. *Advanced Management Journal, 72*, 33-44.

Pawlak, Z. (2002). Rough set theory and its applications. *Journal of Telecommunications and Information Technology, 3*, 7-10.

Perry, R., & Guthrie, J. (2000). Intellectual capital literature review. *Measurement, reporting and management. Bradford,* 1 (2), 155.

Pillania, K. R. (2009). Demystifying knowledge management. *Business Strategy Series*, 10, 96-99.

Plessis, M., & Boon, J. A. (2004). Knowledge management in e-business and customer relationship management: South Africa case study finding. *International Journal of Information Management, 24* (10), 73-85.

Polanyi, M. (1966). *The Tacit Dimension*. London: Routledge and Kegan Paul.

Polanyi, M. (1974). *Personal Knowledge: Towards a Post- Critical Philosophy*. The University of Chicago Press.

Polanyi, M. (1996). *The Tacit Dimension*. The University of Chicago Press, Chicago and London.

Ponis, S. T., & Koronis, E. (2012). Supply Chain Resilience: Definition Of Concepts And Its Formative Elements. *The Journal of Applied Business Research, 28* (5), 921.

Popovic, A., Hackey, R., Coelho, P. S., & Jaklic, J. (2012). Toward business intelligence systems success: Effects of maturity and culture on analytical decision making. *Decision Support Systems, 54*, 729-739.

Prajogo, D. J., & McDermott, C. M. (2005). The relationship between total quality management practices and organizational culture. *International Journal of Operations & Production Management, 25* (11), 1101-1122.

Probst, G., Raub, S., & Romhardd, K. (2000). *Managing Knowledge: Building Blocks for Success*. John Wiley & Sons, Ltd., Baffins Lane, Chichester, UK.

Provost, F. & Fawcett, T. (2013). *Data Science for Business*. O'Reily Media, Inc. Sebastopol, CA. First Edition.

Rabinski, J. S. (2003). Primary and Secondary Data: Concepts, Concerns, Errors and Issues. *The Appraisal Journal*, 71 (1), 43-55.

Raisch, S., Birkinshaw, J., Probst, G., Tushman, M. L. (2009). Organizational Ambidexterity: Balancing Exploitation and Exploration for Sustained Performance. *Organization Science. 20* (4), 685-695.

Rajasekar, S., Philominathan, P., Chinnathambi, V., (2006). Research Methodology. Manuscript.

Ramirez, W. Y. & Steudel J. H. (2008) Measuring knowledge work: the knowledge work quantification framework. *Journal of Intellectual Capital, 9*, 564-584.

Rao, M. (2015). KM Singapore 2015: Twelve tips to unlock the knowledge-ready advantage. *KM World, 24* (10), 3-22.

Rattray, J., Jones, M. C. (2007). Essential elements of questionnaire design and development. *Journal of Clinical Nursing, 16* (2), 234 – 243.

Ray, G., Barney, J. B., & Muhanna, W. A. (2004). Capabilities, business processes, and competitive advantage: choosing the dependent variable in empirical tests of the resource-based view. *Strategic Management Journal, 25* (1), 23-37.

Robb, D. (2000). Building Resilient Organizations. OD Practitioner, 32 (3), 27-32.

Rossi, P. H., Wright, J. D., & Anderson, A. B. (2013). *Handbook of Survey Research. Quantitative Studies in Social Relations.* Academic Press.

Sallam, R. L., Tapadinhas, J., Parenteau, J., Yuen, D., & Hostmann, B. (2014). Magic Quadrant for Business Intelligence and Analytics Platforms ID:G00257740. Retrieved from http://www.gartner.com/technology/reprints.do?id=1-1QHKSEP&ct=140206&st=sb [Retrieved on Jun. 3, 2015.]

Sambamurthy, V., Bharadwaj, A., & Grover, V. (2003) Shaping Agility through Digital Options: Reconceptualizing the Role of IT in Contemporary Firms. *MIS Quarterly, 27* (2), 237-263.

Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research methods for business students*. 5th ed. Pearson Education Limited: Harlow, Essex, England.

Schianetz, K., Kavanagh, L., & Lockington, D. (2007). The learning tourist destination: The potential of learning organization approach for improving the sustainability of tourism destinations. *Tourism Management, 23* (2), 439-454.

Schlogl, C. (2005). Information and knowledge management: dimensions and approaches. *Information Research, 10* (4).

Senge, P. (1990). *The Fifth Discipline*. New York, NY: Doubleday.

Shannack, R. O. (2009). Measuring Knowledge Management Performance. European *Journal of Scientific Research, 35* (2), 242-253.

Shaw, M. J., Subramaninam, C., Tan, G. W., & Welge, M. E. (2001). Knowledge Management and data mining for marketing. *Decision Support Systems, 31,* 127-137.

Shih, K., Chang, C., & Lin, B. (2010). Assessing knowledge creation and intellectual capital in banking industry. *Journal of Intellectual Capital, 11* (1), 74-89.

Shollo, A., & Galliers, R. (2013). Towards An Understanding Of The Role Of Business Intelligence Systems In Organizational Knowing. *ECIS 2013 Completed Research* Paper 164. [European Conference on Information Systems.]

Shollo, A., & Kautz. (2010). Towards an Understanding of Business Intelligence. *Proceedings of ACIS*, Brisbane.

Skyrme, D. J. (1999). From Measurement Myopia to Knowledge Leadership. *KM Performance Measurement Access*, 28-29. London.

Skyrme, D., & Amidon, D. (1998). New Measures of Success. *Journal of Business Strategy, 19* (1), 20-24.

SMB: Sizing up Small-to-Medium Business (SMB) http://smbresearch.net/sizing-up-smb/ [Retrieved Oct. 23, 2013.]

Smith, P. (2011). Information: a Literature Review. *Georgia International Conference on Information Literacy*, Paper 11. http://digitalcommons.georgiasouthern.edu/cgi/viewcontent.cgi?article=1384&context=gaintlit [Retrieved on Oct. 20, 2016]

Smith, W. K., & Lewis, M. W. (2011). Toward a Theory of Paradox: a Dynamic Equilibrium Model of Organizing. *Academy of Management Review*, *36* (2), 381-403.

Spangler, W. E., May, J. H., & Vargas, G. L. (1999). Choosing Data-Mining Methods for Multiple Classification: Representational and Performance Measurement Implications for Decision Support. *Journal of Management Information Systems/ Summer, 16* (1), 37-62.

Spector, P. E. (1981). *Research Designs*. Thousand Oaks: Sage Publications, Inc.

Spender, J. C. (1998). Pluralist Epistemology and the Knowledge-Based Theory of the Firm. *Organization, 5* (2), 233-256.

SPSS, 2010. CRISP-DM 1.0. ftp://ftp.software.ibm.com/software/analytics/spss/support/Modeler/Documentation/14/UserManual/CRISP-DM.pdf. [Retrieved on Aug 24, 2014.]

Stancu, A. M., Ramona, Apetrei, M. C. (2013). Data Mining – Trends and Challenges. *Knowledge Horizons, 5* (2).

Stankosky & Baldanza. (2001). A Systems Approach To Engineering A KM System. Unpublished manuscript.

Starr, R., Newfrock, J., & Delurey, M. (2003). Enterprise Resilience: Managing Risk in the Networked Economy. *Strategy+Business*, 30.

Stats America. Retrieved from http://www.statsamerica.org/profiles/sip_index.html [Retrieved on Oct. 20, 2015].

Sundstrom, G., & Hollnagel, E. (2006). Learning How To Create Resilience in Business Systems. In E. Hollnagel, D. D. Woods, & N. Leveson, *Resilience Engineering: Concepts and Precepts*. Aldershot, UK: Ashgate.

Spencer, S. (2014). The Go-To Resource for B2B Marketers. Retrieved from http://marketeer.kapost.com/survey-response-rates/. [Retrieved on Nov. 14, 2015.]

Sveiby, E. K. (1996). *What is Knowledge Management?* Retrieved from http://www.sveiby.com/articles/KnowledgeManagement.html [Retrieved on May 28, 2010.]

Sveiby, E. K. (1997). *The New Organizational Wealth: Managing and Measuring Knowledge-Based Assets*. Berrett-Koehler Publishers, Inc.

Szulanski, G. (1996). Exploring internal stickiness: Impediments to the transfer of best practice within the firm. *Strategic Management Journal, 17* (10), 27-43.

Tamilselvi J., Jebamalar, & Gifta B. C. (2011). Handling Duplicate Data in Data Warehouse for Data Mining. *International Journal of Computer Applications, 15* (4).

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2*, 53-55.

Teece, D. J., & Leih, S. (2016) Uncertainty, Innovation, and Dynamic Capabilities: An Introduction. *California Management Review*, *58* (4), 5-12.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal, 18* (7), 509-533.

The Midmarket Institute. Retrieved from http://www.midmarket.org/user-type/midsize-companies [Retrieved on Oct. 27, 2015.]

Thurow, L. (1996). *The future of Capitalism*. New York, NY: William Morrow & Co.

Tran, H. Q. (2015). Organizational Ambidexterity in Samll Firms: The role of Top Management Team Behavioral Integration and Entrepreneurial Orientation. *Journal of Business & Economic Policy. 2* (4), 31-39.

Tsai, H. (2012). Global data mining: An empirical study of current trends, future forecasts, and technology diffusions. *Expert Systems with Applications, 39* (9), 8172-8181.

Tsai, H. (2013). Knowledge management vs. data mining: Research trend, forecast and citation approach. *Expert Systems with Applications, 40*, 3160 – 3173.

Tseng, S. M. (2008). Knowledge management system performance measure index. *Expert Systems with Applications, 34* (1), 734-745.

Tsoukas, H. (2002). Do we really understand tacit knowledge? Presented to Knowledge Economy and Society Seminar. LSE Department of Information Systems.

Tsumoto, S. (2013). Special issue on challenges in knowledge discovery and data mining. *J. Intell Inf Syst, 41*, 1-4.

Turban, E., Aronson, J. E., Liang, T. P., & Shadra, R. (2007). *Decision Support and Business Intelligence Systems*. Eight Ed. Upper Saddle River, NJ: Pearson Prentice Hall,.

Underwood, J. (2014a). Analyzing Gartner's 2014 Magic Quadrant for BI and Analytics Platforms. http://www.jenunderwood.com/2014/03/16/analyzing-gartners-2014-magic-quadrant-for-bi-and-analytics-platforms/ [Retrieved on Aug. 25, 2014.]

Underwood, J. (2014b). Analyzing Gartner's 2016 Magic Quadrant for BI and Analytics Platforms. http://www.jenunderwood.com/2016/02/09/big-changes-in-gartners-2016-magic-quadrant-for-bi-and-analytics/ [Retrieved on Feb. 25, 2016.]

University of Arkansas. Sam M. Walton College of Business. 'Data Mining with SQL Server Data Tools. https://walton.uark.edu/enterprise/Microsoft/DataMining/downloads/Example_SQL_Server_Data_Tools_Data_Mining.pdf [Retrieved on Feb. 29, 2016.]

USA Today: www.usatoday.com/storey/money/business/2013/02/24/medium-size-companies-cnbc/1938679/ [Retrieved on Oct. 21, 2015.]

Vapnik, V. (2000). *Statistical Learning Theory*. New York: Wiley.

Vatafu, R. E. (2011). Knowledge Management – The Key Resource For Become Competitive. *Journal of Knowledge Management, Economics and Information Technology, 1* (2).

Venkatraman, N., & Ramanujam, V. (1986). Measurement of Business Performance in Strategy Research: A Comparison of Approaches. *Academy of Management Review, 11* (4), 801-814.

Venzin, M., Krogh, G., & Roos, J. (1998). Future Research into Knowledge Management. In G. Krogh, J. Roos & D. Kleine, *Knowing In Firms Understanding, Managing and Measuring Knowledge*. Sage Publications, Inc.

Vesely, A. (2003). Neural network in data mining. *AGRIC. ECON. – CZECH, 49* (9), 427-431.

Vestal, W. (2002). Measuring Knowledge Management. American Productivity Quality Center. Retrieved from: http://www.providersedge.com/docs/km_articles/measuring_km.pdf [Retrieved on Apr. 24, 2011.]

Vorakulpipat, C., & Rezgui, Y. (2008). Value creation: the future of knowledge management. *The Knowledge Engineering Review, 23* (3), 283-294.

Walliman, N. (2011). *Your Research Project*. London: Sage Publications, Inc.

Wang, C., & Principe, J. C. (1995). Training neural networks with additive noise in the desired signal*. IEEE Transactions on Neural Networks, 10*, 1511-1517.

Wang, H., & Wang, S. (2008). A knowledge management approach to data mining process for business intelligence. *Industrial Management & Data Systems, 10* (5), 622-634.

Wang, K., Yang, J., Shi, G., & Wang, Q. (2008). An expanded training set based validation method to avid overfitting for neural network classifier. *International Conference on Natural Computation, 3*, 83-87.

Wang, T., Touchman, J. W., & Xue, G. (2004). Applying two-level simulated annealing on Bayesian structure learning to infer genetic networks. *Proceedings of the Computational Systems Bioinformatics Conference, IEEE*, 647-648.

Ward, J. H., Jr. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association, 58*, 236–244.

Watson, H. J. (2009). Tutorial Business Intelligence – Past, Present and Future. *Communications of the Association for Information Systems* (25), 487-510.

Watson, H. J. (2010). Business Analytics Insight: Hype or Here to Stay? *Business Intelligence Journal, 16* (1), 4-8.

Watson, H. J., Abraham, D., Chen, D., Preston, D., & Thomas, D. (2004). Data warehousing ROI: justifying and assessing data warehouse. *Business Intelligence Journal*, 6-17.

Weinberger, D. (2010). The Problem with the Data-Information-Knowledge-Wisdom Hierarchy. *Harvard Business Review*. Retrieved from https://hbr.org/2010/02/data-is-to-info-as-info-is-not [Retrieved on Jul. 6, 2013.]

Wenger, E. (1998). *Communities of Practice: Learning, Meaning, and Identity.* Cambridge.

West III, P. G., & Noel, T. W. (2009). The Impact of Knowledge Resources on New Venture Performance. *Journal of Small Business Management, 47* (10), 1-22.

Wiig, K. M. (1997). Integrating intellectual capital and knowledge management. *Long Range Planning 30* (1), 399-405.

Wilson, T. D. (2002). The nonsense of 'knowledge management'. *Information Research, 8* (1).

Witten, H. I., Frank, E., & Hall, M. A. (2011). D*ata Mining: Practical Machine Learning Tools and Techniques* (Third Edition). Burlington MA, USA: Morgan Kaufmann Publishers.

Wright, S., Eid, E.R., & Fleisher, C.S. (2009) Competitive Intelligence in Practice: Empirical Evidence from the UK Retail Banking Sector. *Journal of Marketing Management, 25* (9-10), 941-964.

Wu, W., Lee, Y., Tseng, M., & Chiang, Y. (2010). Data mining for exploring hidden patterns between KM and its performance. *Knowledge-Based Systems, 23*, 397-401.

Yang, C.Y., & Cai, W. (2007). *Extension Engineering*. Beijing: Science Press.

Yli-Renko, H., Autio, E., & Sapienza, H. J. (2001). Social Capital, Knowledge Acquisition, And Knowledge Exploitation In Young Technology-Based Firms. *Strategic Management Journal, 22*, 587-613.

Zhang, C., Yu, P. S., & Bell, D. (2010). Introduction to the Domain-Driven Data Mining. *IEEE Transactions on Knowledge and Data Engineering, 22*, (6).

BIBLIOGRAPHY

Abe, H., & Tsumoto, S. (2012). Detection of research trends from bibliographical data. *IJDMMM* 4 (3), 255–266.

Alvert, K., Borneman, M., Will, M. (2009). Does Intellectual Capital Reporting Matter to Financial Analysts? *Journal of Intellectual Capital,* 10 (3), 354-368.

Baicoianu, A., Dumitrescu, S. (2010). Data Mining Meets Economic Analysis Opportunities and Challenges. *Bulletin in the Transilvania Univeristy of Brasov.* 3 (52).

Bala, L. (2010). Role of Knowledge Management in the Global Business Order. *Economic Challenger,* Jan. 2010.

Berkes, F. (2005). Understanding uncertainty and reducing vulnerability: lessons from resilience thinking. *Nat Hazards,* 41, 283-295.

Boros, E., Crama, Y. (2009). Logical Analysis of Data: Classification with Justification. *In DIMACS Technical Report* , 2009-02.

Bramer, M. A. (2007). *Principles of Data Mining.* London: Springer .

Brown, J. S., Duguid, P. (1991). Organizational Learning and Communities-of-Practice: Toward a Unified View of Working Learning and Innovation. *Organization Science,* 2 (1).

Corso, M., Giacobbe, A., Martini, A. (2009). Designing and managing business communities of practice. *Journal of Knowledge Management, 13* (3), 73-89.

Cross, R., Parker, A., Prusak, L., Borgatti, S. P. (2001). Supporting Knowledge Creation and Sharing in Social Networks. *Organizational Dynamics. 30* (2), 100-120.

Davenport, T. H. (2005). *Thinking for a Living.* Harvard Business School Press.

Dixon, N. M., (2000). *Common Knowledge: How Companies Thrive by Sharing What They Know.* Boston: Harvard Business School Press.

Drucker, Peter F. (1999). *Management Challenges for the 21st Century.* Woburn, MA: Butterworth-Heinemann.

Eccles, R. G. (1991). The Performance Measurement Manifesto. *Harvard Business Review.*

Edvinsson, L. (2002). *Corporate Longitude*, Bookhouse, Stockholm.

Gherardi, S., Nicolini, D., Odella, F. (1998). Toward a Social Understanding of How People Learn in Organizations. *Management Learning, 29* (3), 273-297.

Hemmasi, M., Csanda, C. M. (2009). The Effectiveness of Communities of Practice: An Empirical Study. *Journal of Managerial Issues. 21* (2), 262-279.

Irick, M. L. (2007). Managing Tacit Knowledge in Organization. *Journal of Knowledge Management Practice. 8* (3).

Kimball, R., Ross, M. (2013). *The Data Warehouse Toolkit.* (Third Edition). Wiley.

Knight, B., Knight, K., Jorgenses, A., LeBlanc, P., Davis, M. (2010). *Knight's Microsoft Business Intelligence.* Wrox.

Krause, U. M., Stark, R. (2010). Reflection in Example-and-problem-based learning: effects of reflection prompts, feedback and cooperative learning. *Evaluation & Research in Education. 23* (4), 255-272.

Jhunjhunwala, S. (2009). Monitoring and measuring intangibles using value maps: some examples. *Journal of Intellectual Capital, 10* (2), 211.

Larson, B. (2017). *Delivering Business Intelligence with Microsoft SQL Server 2016* (Fourth Edition). McGraw-Hill Companies.

Liu Chung-Chu. (2010). Prioritizing Enterprise Environment Management Indicators by Intellectual Capital Perspective. *Journal of International Management Studies. 5* (2), 110-117.

Loye, R. (2008). Requirement for knowledge management: business driving information technology. *Journal of Knowledge Management, 12* (3), 156-168.

Mandari, P., Barskar, R. (2013). A Survey of Conceptual Data Mining and Applications. *International Journal of Computer Science and Information Security,* 11 (5), 16.

Massey, A. P., Montoya-Weiss, M. M., O'Driscoll, T. M. (2002). Performance-Centered Design of Knowledge-Intensive Processes. *Journal of Management Information Systems. 18* (4), 37-58.

Masud, K., Rashedur, M. R. (2013). Decision Tree and Naïve Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing. *Journal of Software Engineering and Applications, 6*, pp. 196-206.

McDargh, E. (2003). Mastering Resilience Skills for Off-the-Charts Results. *Management Quarterly, 44* (1), 2.

McElroy, M.W. (2000). Using Knowledge Management to sustain innovation. *Knowledge Management Review*, *3* (4), 34-37.

McElroy, M. W. (2000). Integrating complexity theory, knowledge management and organizational learning. *Journal of Knowledge Management, 4* (3), 195.

McInerney, C. (2002). Knowledge Management and the Dynamic Nature of Knowledge. *Journal of the American Society for Information Science and Technology, 53* (12), 1009-1018.

Narvekar, R.S., Jain, K. (2006). A new framework to understand the technological innovation process. *Journal of Intellectual Capital*, *7* (2).

O'Neal M. R. (1999). Measuring Resilience. *Annual Meeting of the Mid-South Educational Research Association* (Point Clear, AL.)

Rad, R. (2014). *Microsoft SQL Server 2014 Business Intelligence Development.* Packt.

Raeder, T., Chawla, N. V. (2011) *Market Basket Analysis with Networks*. Springer.

Resnick, M. L., Mejia, A. (2007). Communities of Practice: Knowledge Management for the Global Organization. *Proceedings of the 2007 Industrial Engineering Research Conference.*

Robinson, H. S., Carrillo, P. M., Al-Ghassani, A. M. (2006). STEPS: a knowledge management maturity roadmap for corporate sustainability. *Business Process Management Journal. 12* (6), 793.

Roos, G., Roos, J. (1997). Measuring your Company's Intellectual Performance. *Long Range Planning, 30* (3), 413-426.

Rothwell, R. (1994). Towards the Fifth-generation Innovation Process. *International Marketing Review, 11* (1), pg. 7.

Ruderman, M. N., Ernst, C. (2004). Finding Yourself How Social Identity Affects Leadership. *Leadership in Action, 24* (3), 3-7.

Serrat, O. (2009). Social Network Analysis. *Asian Development Bank.* 28.

Serrat, O. (2009). A Primer on Organizational Learning. *Asian Development Bank.* 69.

Spangler, W. E, May, J. H., Vargas, L. G. (1999). Choosing Data-Mining Methods for Multiple Classification: Representational and Performance Measurement Implications for Decision Support. *Journal of Management Information Systems, 16* (1), 37-62.

Stubbs, E. (2011). *The Value of Business Analytics.* Wiley.

Tunguz, T., Bien, F. (2016). *Winning with Data.* Wiley.

Urbancic , T., Skrjanc , M. and Flach , P. (2002). Web-based analysis of data mining and decision support education . *AI Communications, 15* (4), 199 – 204.

Villalonga., B. (2004). Intangible resources, Tobin's q, and sustainability of performance differences. *Journal of Economic Behavior & Organization. 54*, 205-230.

Von Krogh, G. K. and Roos, J., (1996). *Managing Knowledge: Perspectives on Cooperation and Competition*. London: Sage Publications, Inc.

Webb, C., Ferrari, A., Russo, M. (2014). *Expert Cube Development with SSAS Multidimensional Models.* Packt.

Wiggins, R. R., & Ruefli, T. W. (2002). Sustained competitive advantage: Temporal dynamics and the incidence and persistence of superior economic performance. *Organization Science, 13*, 82–105.

Wilson, F. (2008). Resilience, the new competitive advantage. *Profit, 26* (6), 44-48.

Winter, S.G., (1987)*. Knowledge and competence as strategic assets*. In: Teece, D.J. (Ed.), The Competitive Challenge. Ballinger, Cambridge, MA.

Zboralski, K. (2009). Antecedents of knowledge sharing in communities of practice. *Journal of Knowledge Management, 13* (3), 90-101.

Zhang, H. (2004). The optimality of Naive Bayes. *The 17th International FLAIRS Conference, Florida Artificial Intelligence Research Symposium*, Miami Beach, FL

APPENDIX I: The Questionnaire

Dear Executive,

I recognize that the demands on your time are enormous, so my appreciation for your participation in this academic research project cannot be overstated. I am truly grateful for your time, and I hope to provide something of value for your organization in return for approximately 30 minutes of your time. This questionnaire is a chance for you to state your opinions for the benefit of mid-size businesses based in Midwest as well as the benefits of society.

My name is Michael Frelas. I am doctoral researcher studying the impact of knowledge management on organizational resilience within mid-size companies operating in the Midwest area of the US. [In short, I am trying to determine how successful companies are using and managing knowledge so that they stay at the top of their game.] While this work is conducted at a Scottish University (Robert Gordon University) I am a US citizen residing in the NW suburbs of Chicago.

By completing this questionnaire, you will be contributing to research in the field. Your input is of great value to this work and is greatly appreciated. In return for your time devoted to answering this questionnaire you will be provided (free of charge) with a feedback on your organization's performance vs. other participating companies. A free copy of my doctoral thesis will also be available. Please indicate if you wish to receive a copy at the end of this questionnaire. Please note, any responses you provide will be treated confidentially, and your anonymity will be preserved.

Finally, your reflection on the questionnaire's weak points (see the very last page) would be of extreme value to me and to this research. Please share your observations and/or opinions.

Once again, I would like to thank you for your time.

Please do not hesitate to contact me or my research supervisor if you have any questions or comments related to this research.

Best regards,

Michael Frelas, Doctoral Candidate

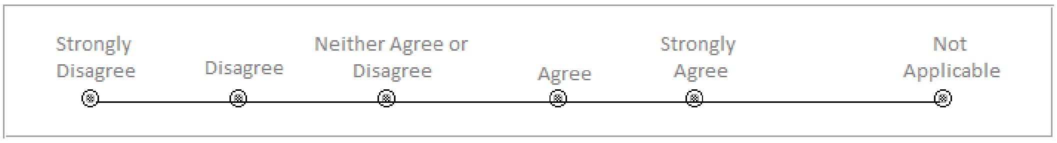
m.frelas@rgu.ac.uk

Cell phone #: (773) 505-8377

Research Supervisor: Dr. Simon Burnett s.burnett@rgu.ac.uk

<u>Questionnaire</u> *Please state your position:* ____

For each question please circle one point on the scale which you feel most closely represents your opinion. An example of the scale used in this questionnaire is provided below:

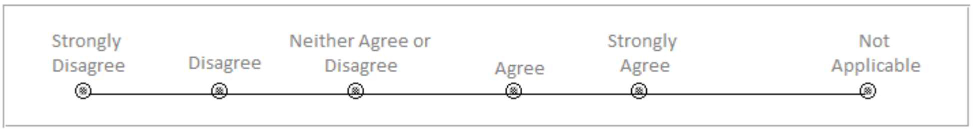


For each question, ‘company’, ‘organization’ and ‘firm’ are treated synonymously and refer to the company that currently employs you.

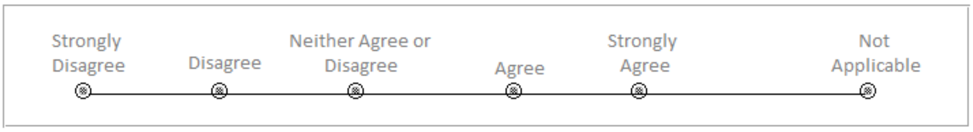
The questionnaire contains 84 questions, and it should take approximately 30 minutes to complete.

The following questions relate to the knowledge creation/acquisition and knowledge exploration in your organization. Answers to these questions help to understand any gaps in knowledge or knowledge-related processes that can provide insight into the competitiveness of your organization.

1. In the last two years my organization has identified and evaluated gaps in its knowledge that need to be filled in order to compete successfully.



2. As a result of identifying and evaluating knowledge gaps (question # 1), my company took corrective actions.



3. As a result of an evaluation of knowledge in my company (question # 1) I felt satisfied that no further action was needed to close the knowledge gaps.

Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

4. The employees at my company are formally encouraged to take the time during their work day to think about better ways of performing their jobs and/or about enhancements of the company's products or services.

Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

5. The company provides physical facilities (conference rooms, break rooms, etc.) for employees to exchange ideas among themselves.

Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

6. Employees' suggestions about improvements to their jobs or work processes and/or product offerings are recorded, stored and are easily accessible by other employees.

Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

7. Employees are allowed to experiment with their ideas to determine their viability.

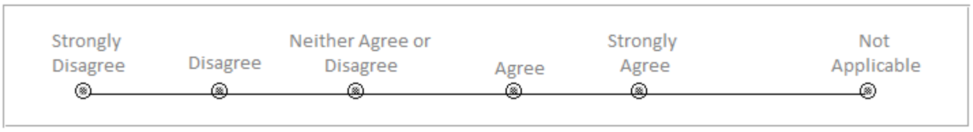
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

8. My company uses one or more of the following (or similar sources) to gain insights:
   * Outsider's market data
   * Comparative data
   * Customer feedback

Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

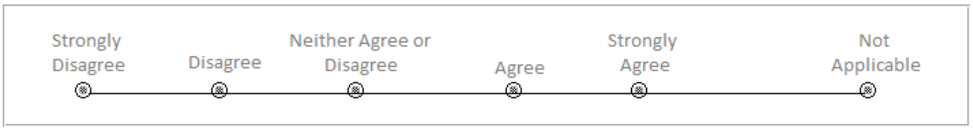
The following questions are related to the exploitation of existing knowledge in your organization. Answers to these questions help to understand adaptive learning taking place at your organization.

9. Prior to a major event/project I typically access company databases, intranets and/or other internal electronic sources of information for reference.



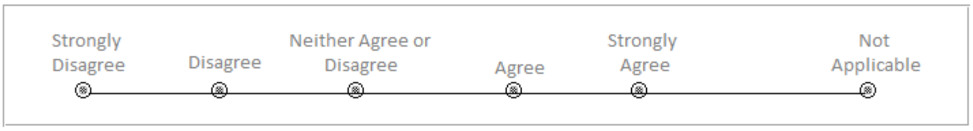
○ Strongly Disagree ○ Disagree ○ Neither Agree or Disagree ○ Agree ○ Strongly Agree ○ Not Applicable

10. It is common for me to simulate a major event/project to walk through possible scenarios.



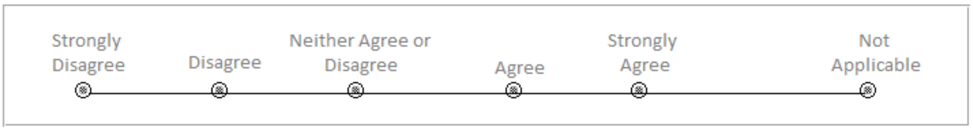
○ Strongly Disagree ○ Disagree ○ Neither Agree or Disagree ○ Agree ○ Strongly Agree ○ Not Applicable

11. It is common before a major event/project that key participants consult with colleagues who have experienced similar events/projects in the past.



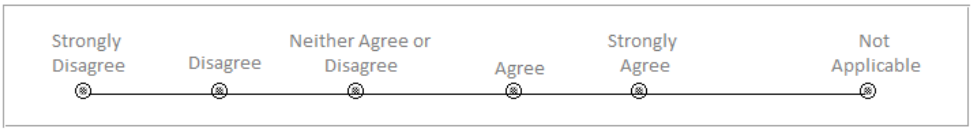
○ Strongly Disagree ○ Disagree ○ Neither Agree or Disagree ○ Agree ○ Strongly Agree ○ Not Applicable

12. It is common practice at my organization to record and electronically store key aspects of an event/project as they occur, or shortly afterwards.



○ Strongly Disagree ○ Disagree ○ Neither Agree or Disagree ○ Agree ○ Strongly Agree ○ Not Applicable

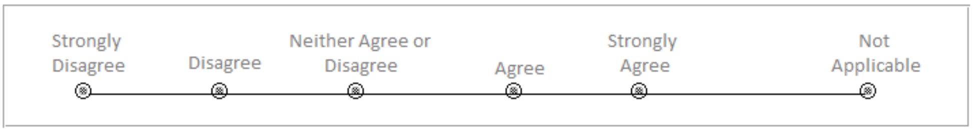
13. It is common practice at my organization to reflect on an entire major project/event after such a project/event has been completed.



○ Strongly Disagree ○ Disagree ○ Neither Agree or Disagree ○ Agree ○ Strongly Agree ○ Not Applicable

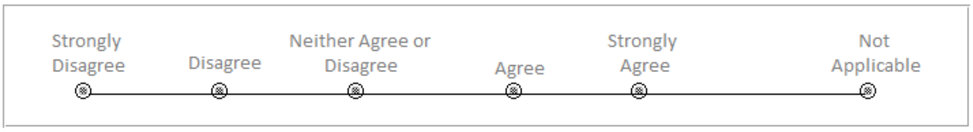
The following questions are related to the decision making and decision alignment in your organization. Answers to these questions help to understand the effects of accessing and integrating diverse information and knowledge as a part of decision making process.

14. My organization forms alliances and joint ventures.



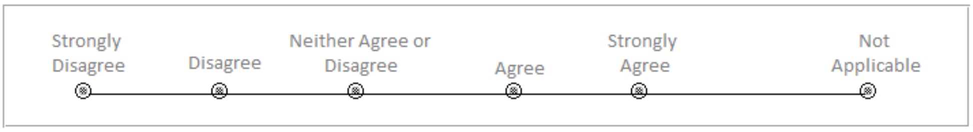
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

15. My organization is involved in co-operative product/service development initiatives.



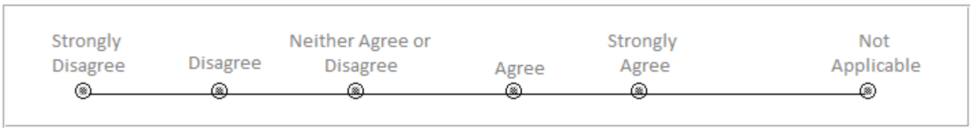
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

16. My organization is actively involved in partnerships with its suppliers.



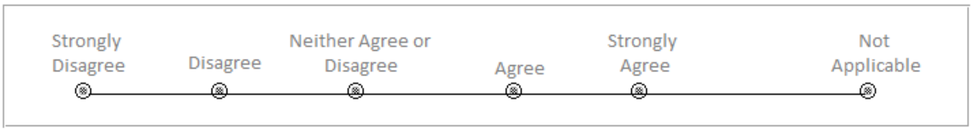
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

17. My organization participates in industry standards initiatives.



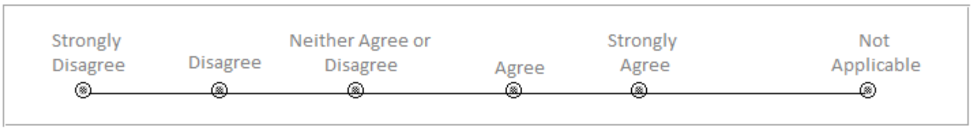
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

18. I view membership of professional organizations as excellent learning opportunities.



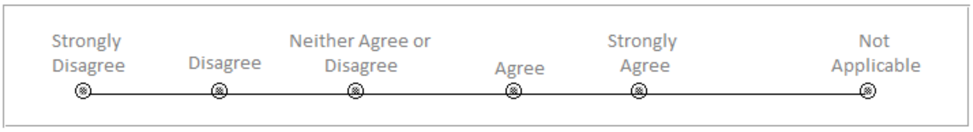
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

19. My organization values engagements in local organizations, like local Chambers of Commerce.



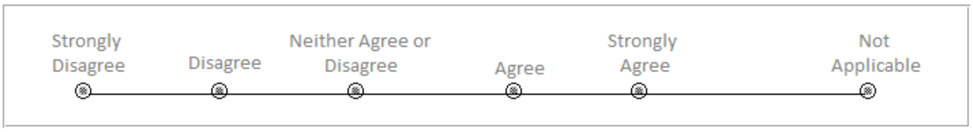
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

20. My organization actively supports customer communities.



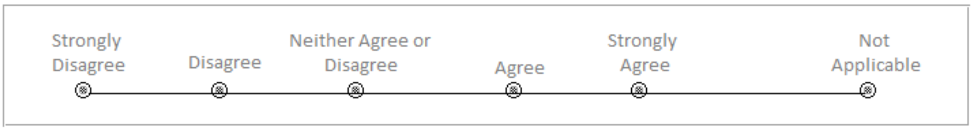
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

21. My organization currently sponsors university/academic research.



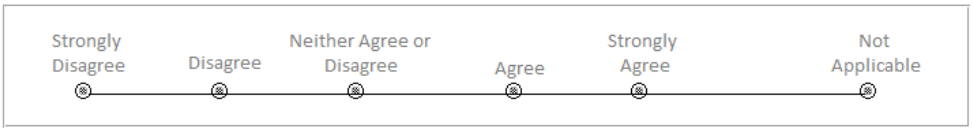
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

22. My organization uses an in-house competitive intelligence system.



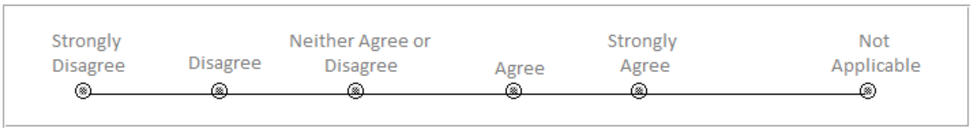
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

23. My organization uses a CRM (Customer Relationship Management) system and views CRM as a highly strategic tool.



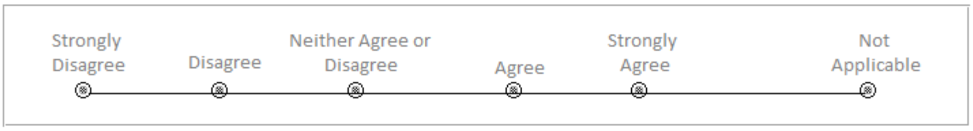
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

24. My organization provides work conditions that encourage individuals to be attentive to their work, and to the needs of colleagues.



Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

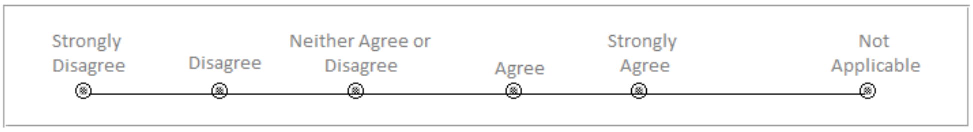
25. My organization sets boundaries for decision-making based on intrinsic factors such as values and ethics.



Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

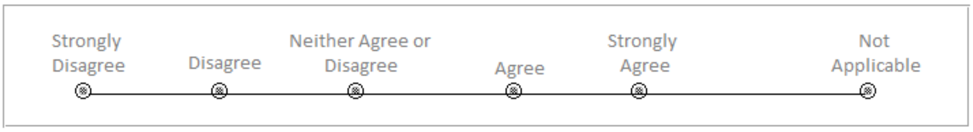
The following questions are related to individual and organizational learning. Answers to these questions help to evaluate the level of learning in your organization.

26. My organization regularly offers formal training to its employees.



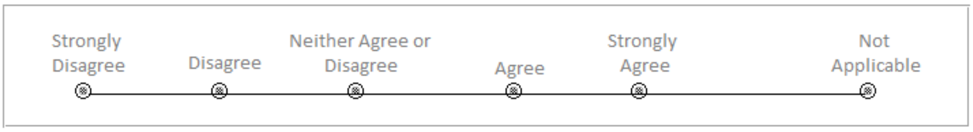
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

27. My organization offers ‘on the job’ learning opportunities (such as apprenticeship, mentoring, etc.).



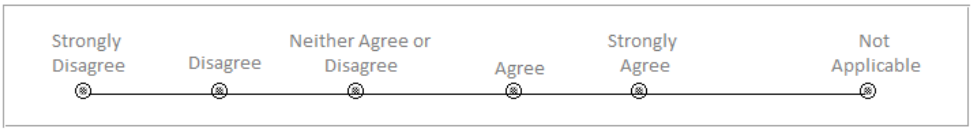
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

28. My organization reimburses employees for continuing education/formal education classes.



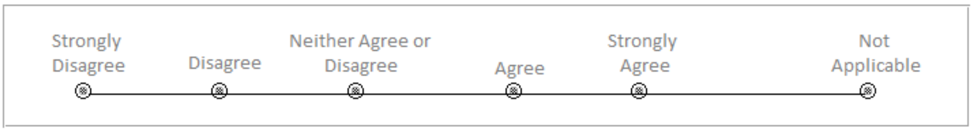
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

29. My company makes organizational learning a priority.



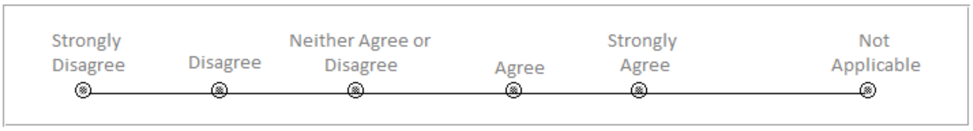
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

30. My organization has a formal process (or processes) for capturing lessons learned.

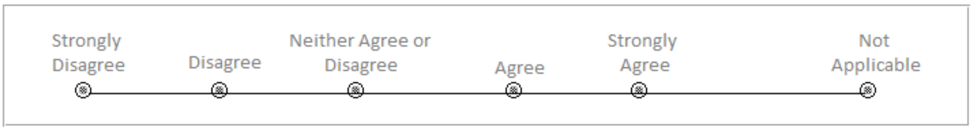


Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

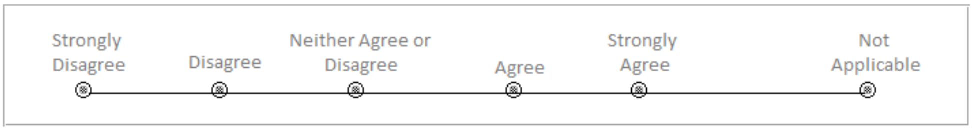
31. My organization embraces and provides venues for verbal exchange of experience and knowledge within organized groups (sometimes refer to as communities of practice).



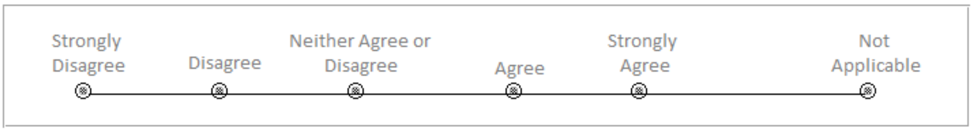
32. It is common in my organization for teams/groups to meet in off-site locations in order to participate in work-related or leisure activities.



33. My organization has an in-house (Intranet) portal for sharing information with employees.

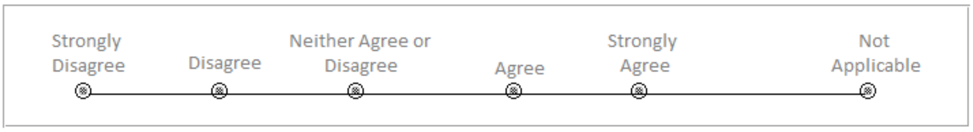


34. My organization has an in-house business intelligence system for data mining purposes (detecting and predicting sales patterns, grouping customers based on some characteristic, etc.).

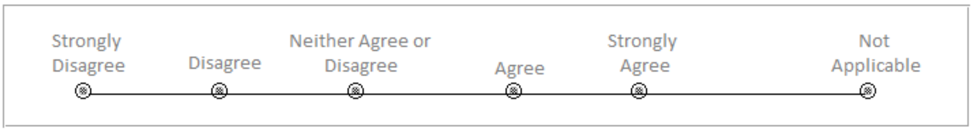


The following questions are related to connecting intra-organizational activities with those activities occurring outside of organizational boundaries. Answers to these questions help to understand how knowledge can be a source of internal and external influence, which knowledge to protect and which new ideas to absorb.

35. Employees at my organization have common, shared beliefs and values.

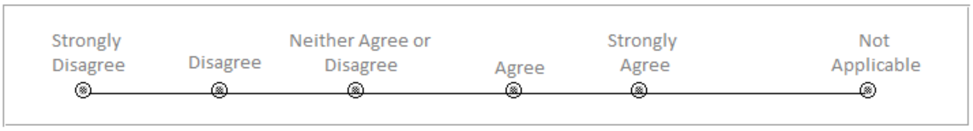


36. Employees at my organization feel empowered.



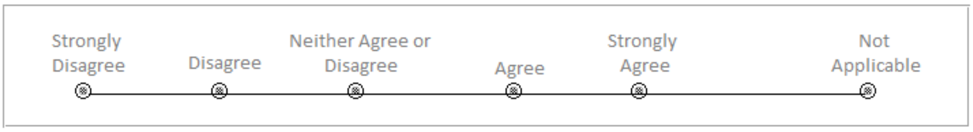
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

37. I feel confident about the competitive position of my firm at the moment.



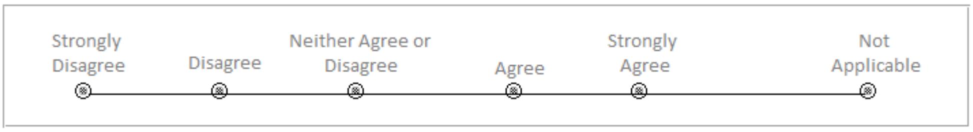
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

38. My organization would be able to identify and quickly act on breakthrough information.



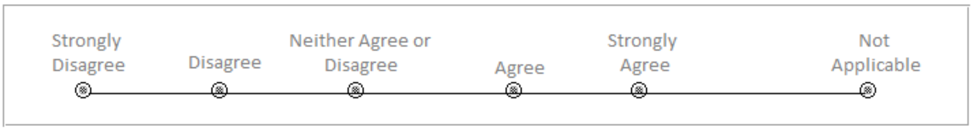
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

39. My organization seeks to build strong relationships with customers and use their feedback so that is can be used strategically.



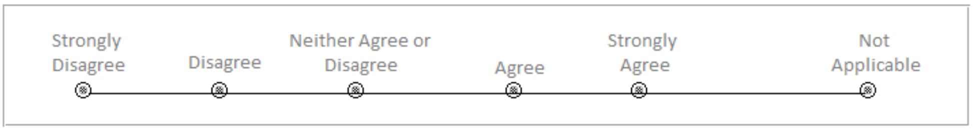
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

40. My organization systematically evaluates political, economic, social, technological and environment changes, and make changes to its strategy as result of this evaluation.



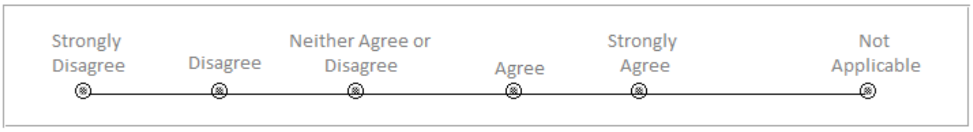
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

41. At least annually the company evaluates its competitors and their activities.



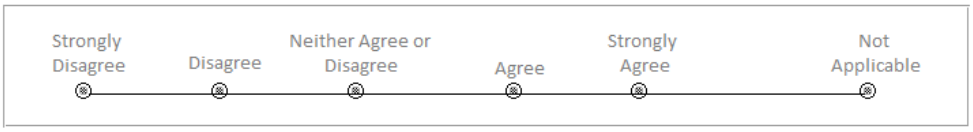
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

42. We actively engage in educating customers, or the general public about the firm's products or services and the direct or indirect benefits of their use.



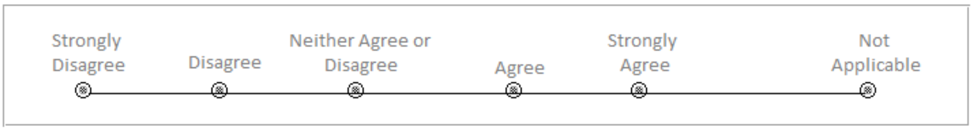
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

43. My organization actively engages in a buying coalition or work on educating its suppliers.



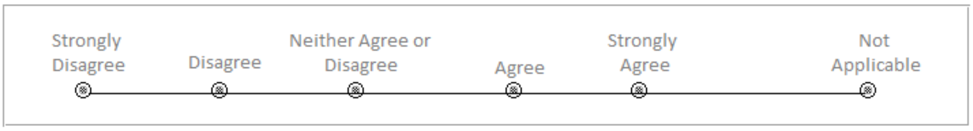
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

44. The company actively seeks to engage in shared business activities with firms in complimentary industries.



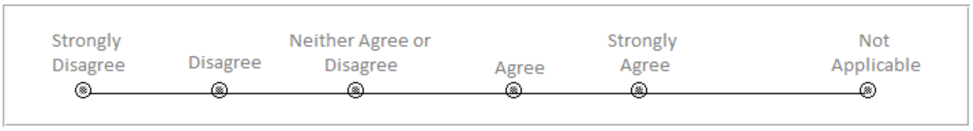
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

45. The company has considered sharing resources with its competitors in non-competitive areas.



Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

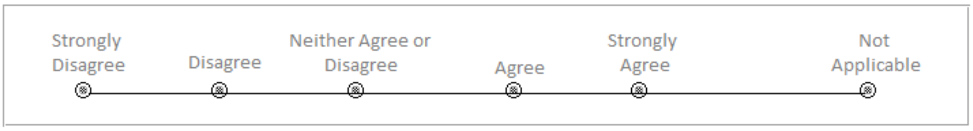
46. My organization has a process (or processes) in place to protect valuable organizational knowledge.



Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

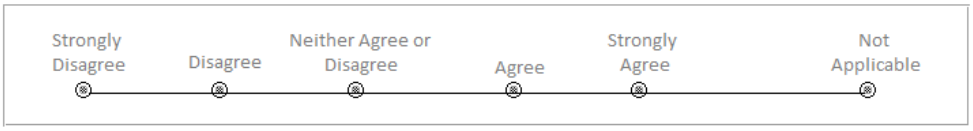
The following questions address the business links that your organization currently has. Answers to these questions help to understand the connections between your organization and external partners as well as external resources.

47. In the last 3 years we have made closer relationships with customers, suppliers and other external partners.



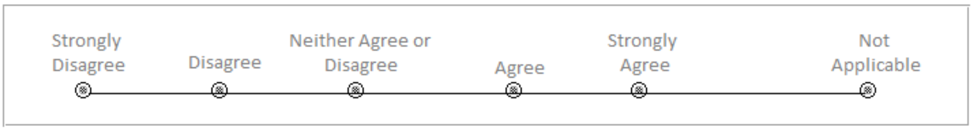
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
⊛	⊛	⊛	⊛	⊛	⊛

48. My organization has a designated person responsible for making closer relationships with stakeholders.



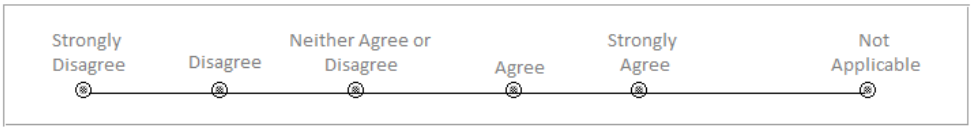
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
⊛	⊛	⊛	⊛	⊛	⊛

49. My organization actively manages problems arising in relationships with stakeholders.



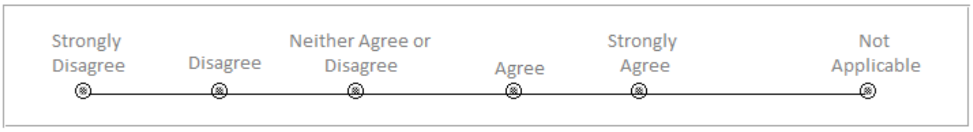
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
⊛	⊛	⊛	⊛	⊛	⊛

50. My organization actively manages its outsourcing relationships.



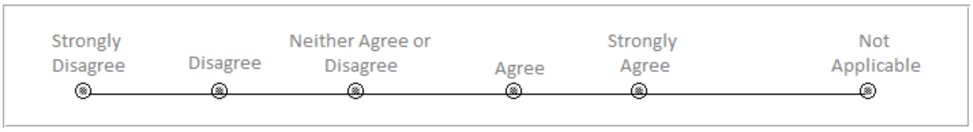
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
⊛	⊛	⊛	⊛	⊛	⊛

51. My organization actively monitors the performance of its relationships with stakeholders, and identifies opportunities to generate more value from building closer connections.



Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
⊛	⊛	⊛	⊛	⊛	⊛

52. We are building leadership expertise in managing loose external relationships.

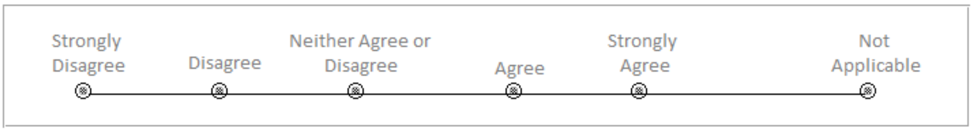


Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
⊛	⊛	⊛	⊛	⊛	⊛

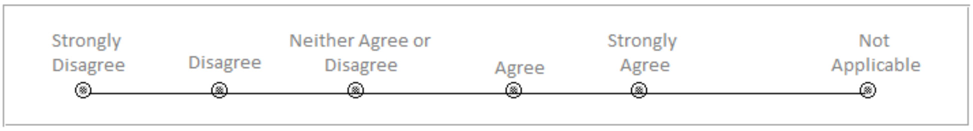
The following questions evaluate your firm's current (as well as expected future) performance.

The answers to these questions provide insight into the evaluation of performance of intellectual capital, the communication of such value to external investors, as well as to the valuation of intellectual capital by your organization.

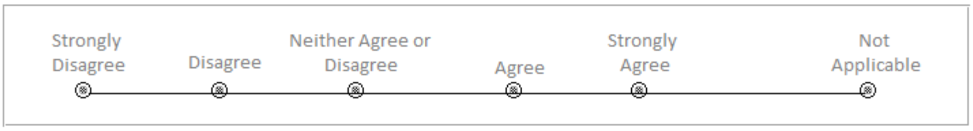
53. My organization currently monitors individual employees' competence and organizational knowledge assets in relation to the demands of its competitive environment.



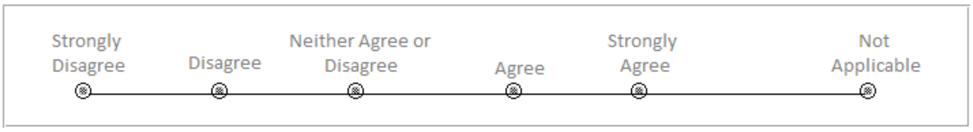
54. My organization tracks employees' satisfaction as it is likely to result in the willingness of individuals to apply their competence to improve organizational performance.



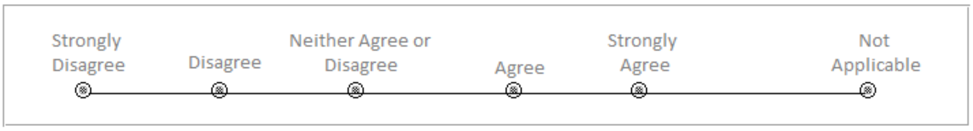
55. My organization has an inventory of employees' competencies.



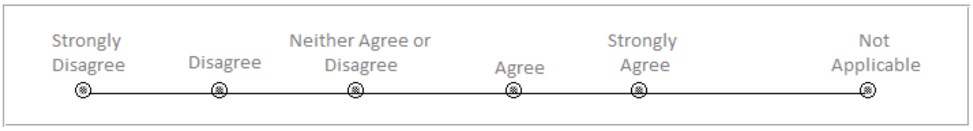
56. My organization monitors the effectiveness of performance management as well as the reward system in generating valuable outcomes through the use of its knowledge.



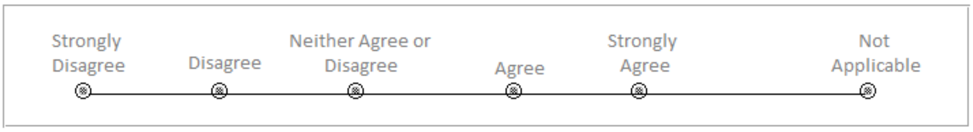
57. My organization tracks financial contributions generated from activities set out to improve products/services/business processes (like suggestions, action reviews, post project reflections, etc.).



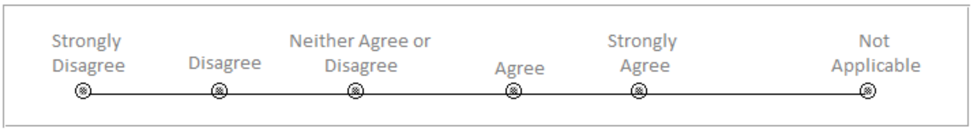
58. In the last 12 months my organization evaluated the contribution of external relationships with its stakeholders to the performance of the company.



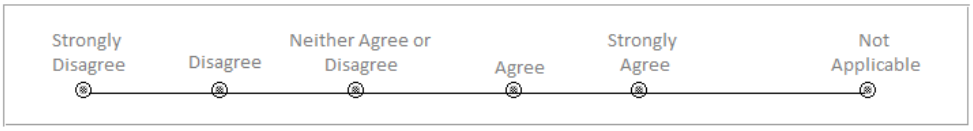
59. We know our company's brand image as perceived by our three best customers.



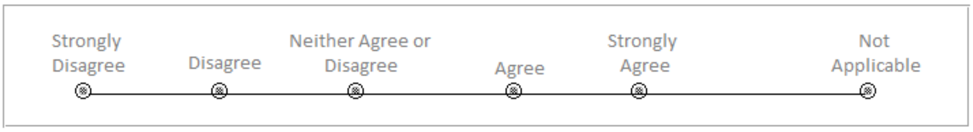
60. In the last 36 months my organization was able to obtain copyrights and/or trademarks.



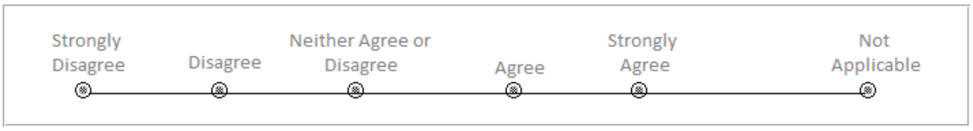
61. I view the collaborative climate of my organization as excellent.



62. The top management in my organization creates conditions for effective collaboration.

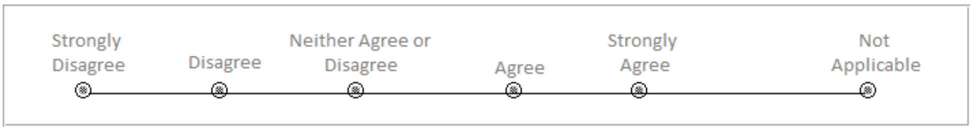


63. In the past 12 months I have received at least one business proposal that would challenge current business strategy and/or business objectives.



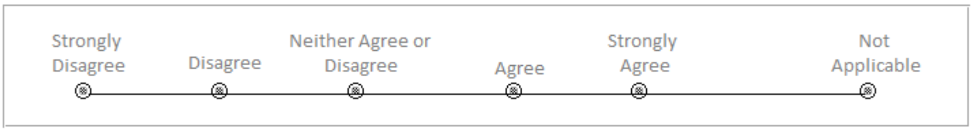
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

64. My organization’s speed of accessing, assimilating and translating information into actionable items is excellent.



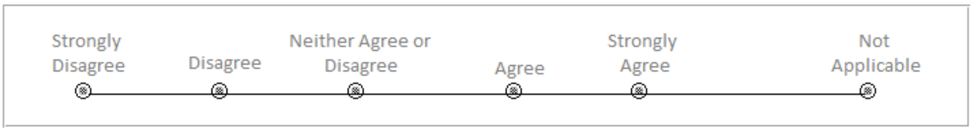
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

65. The ratio of loosely-connected and closely-tied relationships within my organization’s network is about the same (as opposed to more loosely-connected or more tightly-coupled connections).



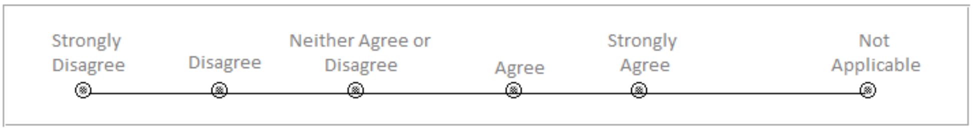
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

66. I am satisfied with the diversity of backgrounds and experience represented across our current workforce.



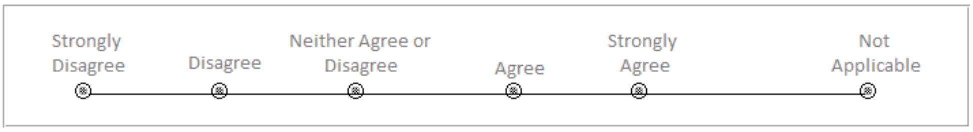
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

67. In the last 12 months my organization validated the usefulness of competitive analysis as well as other external data it uses for its strategy.



Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

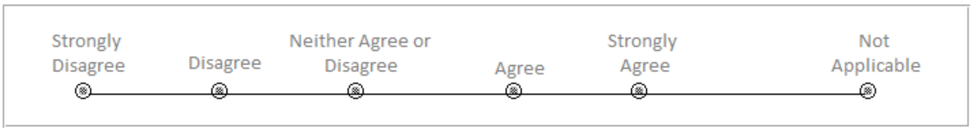
68. When faced with a problem we form a positive and constructive perception of the issue.



Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

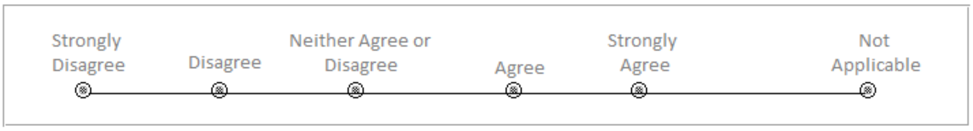
The following questions relate to organizational resilience: the ability of your organization to be successful in spite of adverse business conditions. Answers to these questions will provide insight into your organization's level of resilience.

69. When faced with an important business problem I view the problem as an opportunity rather than a threat.



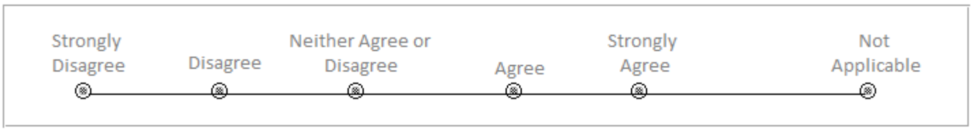
○ Strongly Disagree ○ Disagree ○ Neither Agree or Disagree ○ Agree ○ Strongly Agree ○ Not Applicable

70. When faced with an important business problem I feel that I have necessary external resources available to deal with the problem.



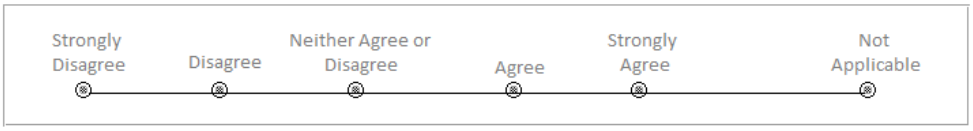
○ Strongly Disagree ○ Disagree ○ Neither Agree or Disagree ○ Agree ○ Strongly Agree ○ Not Applicable

71. I have developed a tolerance for uncertainty.



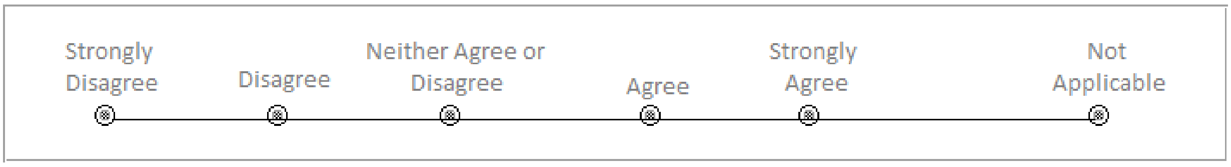
○ Strongly Disagree ○ Disagree ○ Neither Agree or Disagree ○ Agree ○ Strongly Agree ○ Not Applicable

72. My company is entirely free of denial, nostalgia and arrogance when dealing with change occurring because of variations in business conditions, in market or in a political arena.



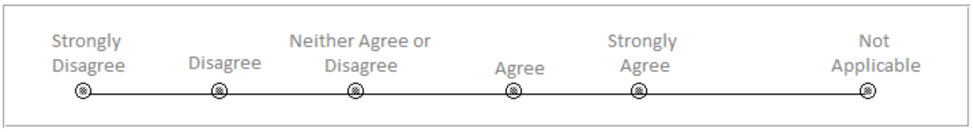
○ Strongly Disagree ○ Disagree ○ Neither Agree or Disagree ○ Agree ○ Strongly Agree ○ Not Applicable

73. When faced with new challenges my company can create a plethora of new options as compelling alternatives to dying strategies.



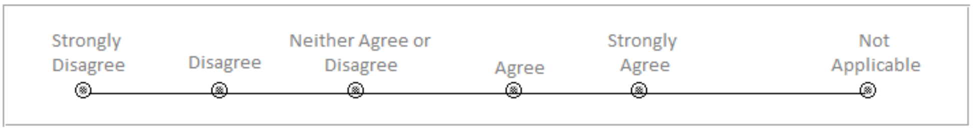
○ Strongly Disagree ○ Disagree ○ Neither Agree or Disagree ○ Agree ○ Strongly Agree ○ Not Applicable

74. We can quickly divert resources from yesterday’s products and programs to those of tomorrow.



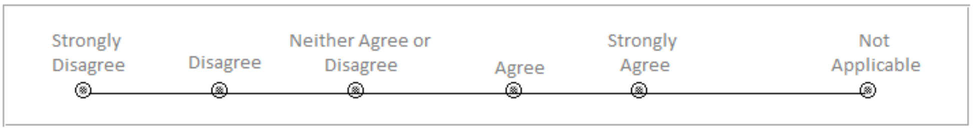
○ Strongly Disagree ○ Disagree ○ Neither Agree or Disagree ○ Agree ○ Strongly Agree ○ Not Applicable

75. My organization sees innovation as more important than optimization of operations.



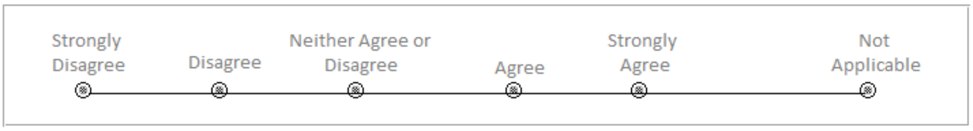
○ Strongly Disagree ○ Disagree ○ Neither Agree or Disagree ○ Agree ○ Strongly Agree ○ Not Applicable

76. Since the beginning of the financial crisis in 2008 we were faced with at least one major business turnaround.



○ Strongly Disagree ○ Disagree ○ Neither Agree or Disagree ○ Agree ○ Strongly Agree ○ Not Applicable

77. My company’s net income has at least improved slightly (if not better) in the last 10 years.



○ Strongly Disagree ○ Disagree ○ Neither Agree or Disagree ○ Agree ○ Strongly Agree ○ Not Applicable

78. My company’s net income has at least improved slightly (if not better) in the last 5 years.

Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

79. My company’s market share has at least improved slightly (if not better) in the last 10 years.

Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

80. My company’s market share has at least improved slightly (if not better) in the last 5 years.

Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

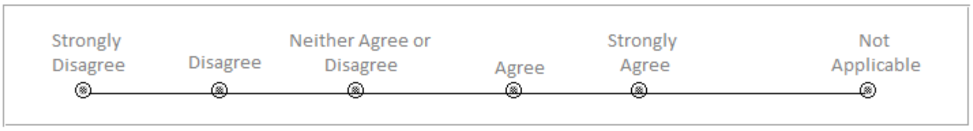
81. My company’s assets have at least improved slightly (if not better) in the last 10 years.

Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

82. My company’s assets have at least improved slightly (if not better) in the last 5 years.

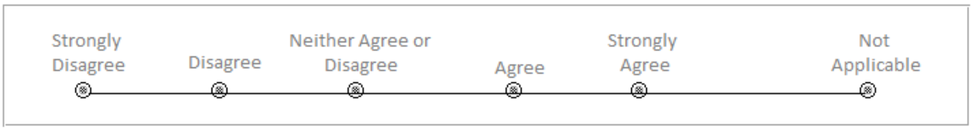
Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

83. My company tends to manage focusing on long-term goals rather than short-term benefits.



Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

84. My organization is capable of innovation and change.



Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	Not Applicable
○	○	○	○	○	○

Please provide an email address where the final report and findings of this research can be sent:

Email: ____

Please indicate if you wish to receive electronic copy of the doctoral thesis: Yes_____, No_____

Thank you very much for your time and your contribution to this research!

To make this questionnaire better please provide your feedback below.

Thank you very much again for your time!

1. Is this research questionnaire manageable in length?
   ____

2. Is this research questionnaire manageable in complexity?
   ____

3. Are the questions clear? If not, which questions need more clarification?

   ____

4. Do you feel that some of the questions were too general? If so, which questions appeared to be too general?
   ____

5. How can this questionnaire be improved?

____

____

____

____

____

____

____

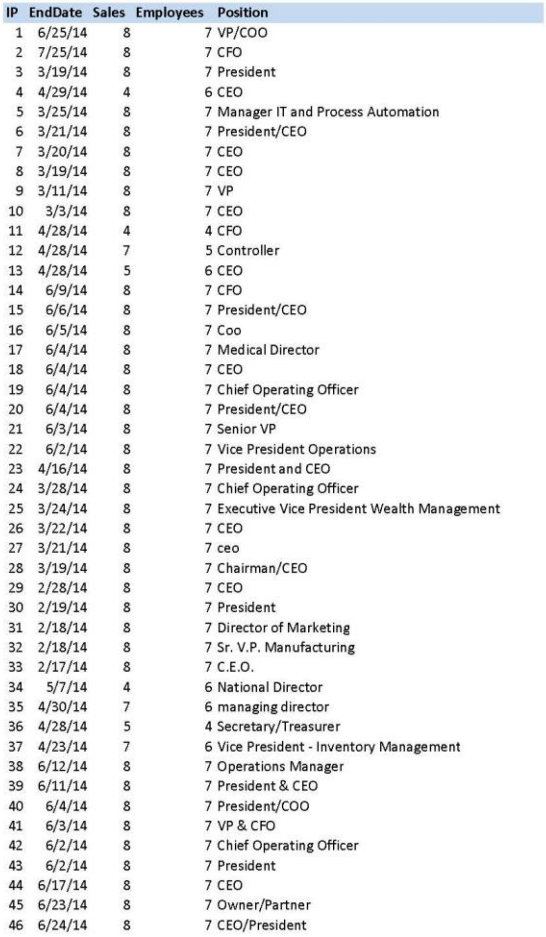
____

____

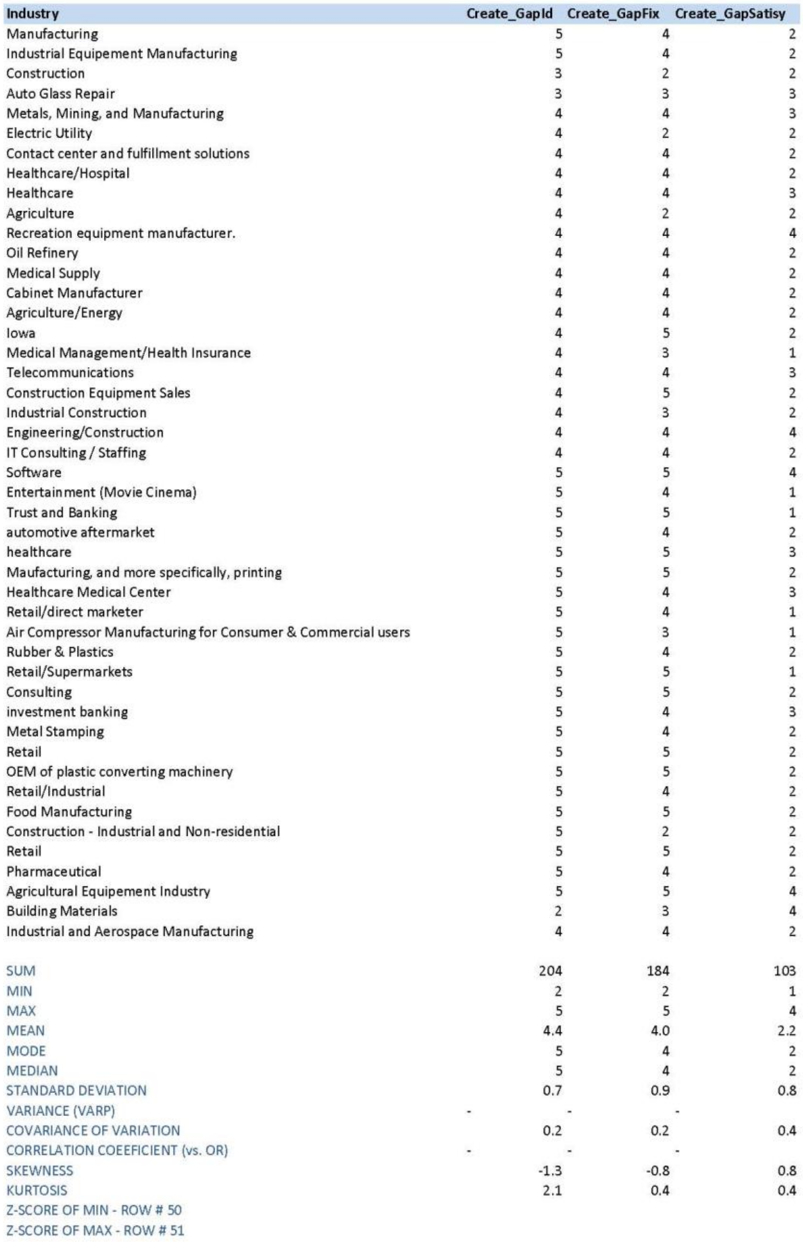
____

____

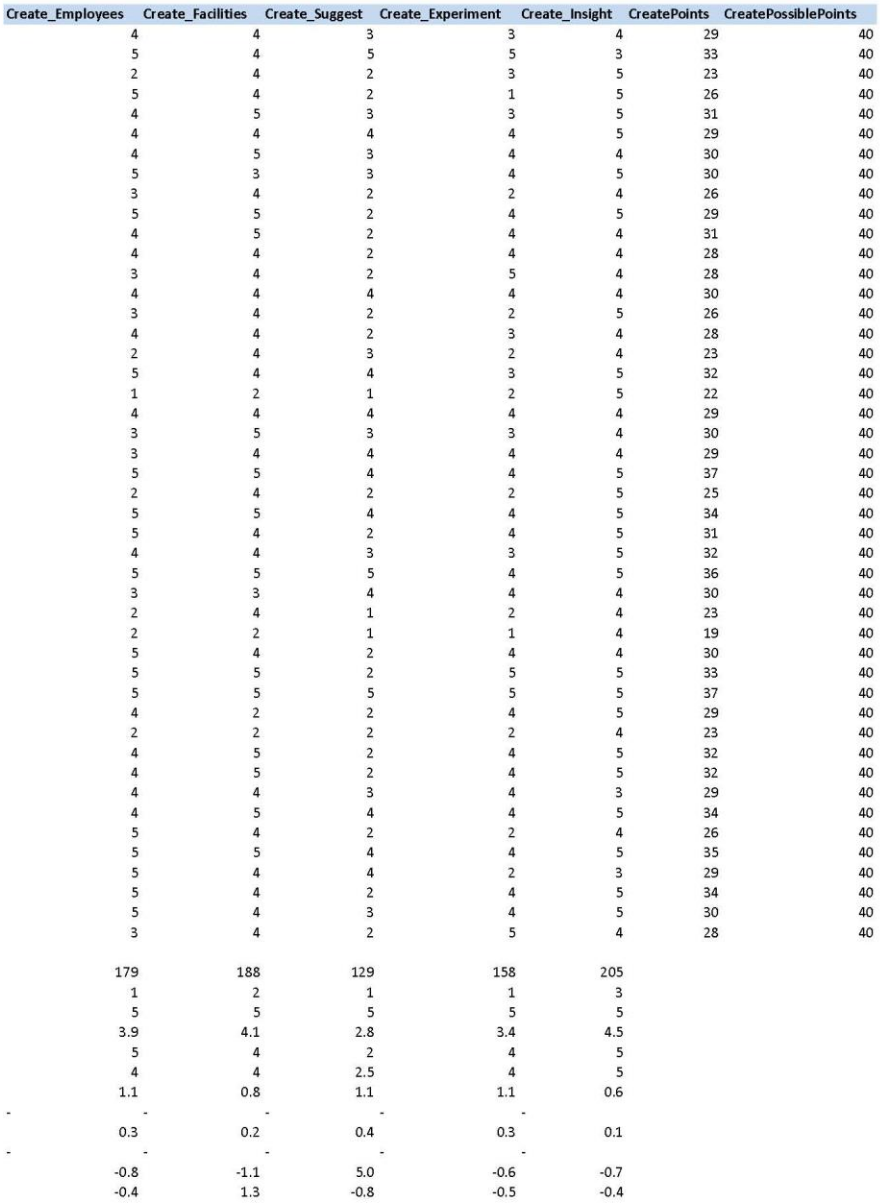
APPENDIX II: Questionnaire Responses



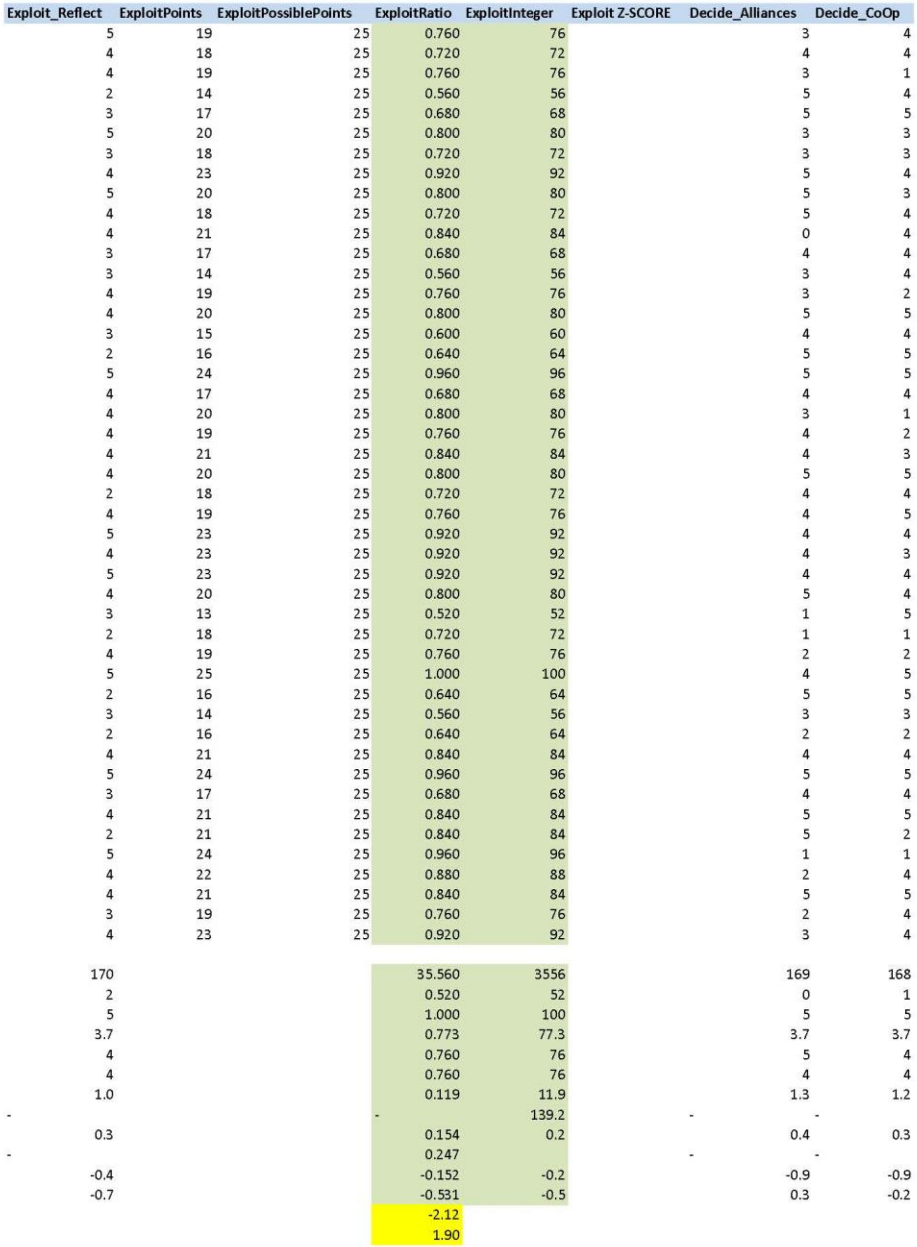
IP	EndDate	Sales	Employees	Position
1	6/25/14	8	7	VP/COO
2	7/25/14	8	7	CFO
3	3/19/14	8	7	President
4	4/29/14	4	6	CEO
5	3/25/14	8	7	Manager IT and Process Automation
6	3/21/14	8	7	President/CEO
7	3/20/14	8	7	CEO
8	3/19/14	8	7	CEO
9	3/11/14	8	7	VP
10	3/3/14	8	7	CEO
11	4/28/14	4	4	CFO
12	4/28/14	7	5	Controller
13	4/28/14	5	6	CEO
14	6/9/14	8	7	CFO
15	6/6/14	8	7	President/CEO
16	6/5/14	8	7	Coo
17	6/4/14	8	7	Medical Director
18	6/4/14	8	7	CEO
19	6/4/14	8	7	Chief Operating Officer
20	6/4/14	8	7	President/CEO
21	6/3/14	8	7	Senior VP
22	6/2/14	8	7	Vice President Operations
23	4/16/14	8	7	President and CEO
24	3/28/14	8	7	Chief Operating Officer
25	3/24/14	8	7	Executive Vice President Wealth Management
26	3/22/14	8	7	CEO
27	3/21/14	8	7	ceo
28	3/19/14	8	7	Chairman/CEO
29	2/28/14	8	7	CEO
30	2/19/14	8	7	President
31	2/18/14	8	7	Director of Marketing
32	2/18/14	8	7	Sr. V.P. Manufacturing
33	2/17/14	8	7	C.E.O.
34	5/7/14	4	6	National Director
35	4/30/14	7	6	managing director
36	4/28/14	5	4	Secretary/Treasurer
37	4/23/14	7	6	Vice President - Inventory Management
38	6/12/14	8	7	Operations Manager
39	6/11/14	8	7	President & CEO
40	6/4/14	8	7	President/COO
41	6/3/14	8	7	VP & CFO
42	6/2/14	8	7	Chief Operating Officer
43	6/2/14	8	7	President
44	6/17/14	8	7	CEO
45	6/23/14	8	7	Owner/Partner
46	6/24/14	8	7	CEO/President



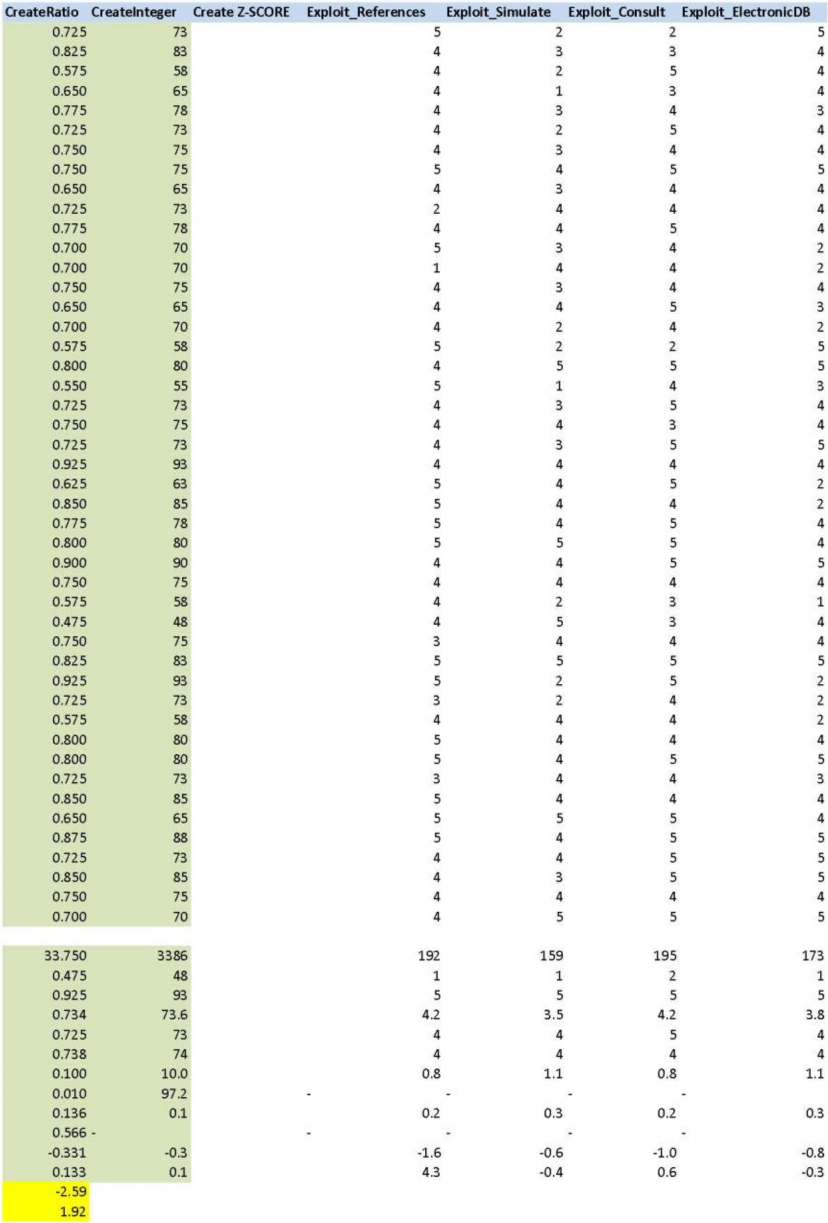
Industry	Create_GapId	Create_GapFix	Create_GapSatisy
Manufacturing	5	4	2
Industrial Equipement Manufacturing	5	4	2
Construction	3	2	2
Auto Glass Repair	3	3	3
Metals, Mining, and Manufacturing	4	4	3
Electric Utility	4	2	2
Contact center and fulfillment solutions	4	4	2
Healthcare/Hospital	4	4	2
Healthcare	4	4	3
Agriculture	4	2	2
Recreation equipment manufacturer.	4	4	4
Oil Refinery	4	4	2
Medical Supply	4	4	2
Cabinet Manufacturer	4	4	2
Agriculture/Energy	4	4	2
Iowa	4	5	2
Medical Management/Health Insurance	4	3	1
Telecommunications	4	4	3
Construction Equipment Sales	4	5	2
Industrial Construction	4	3	2
Engineering/Construction	4	4	4
IT Consulting / Staffing	4	4	2
Software	5	5	4
Entertainment (Movie Cinema)	5	4	1
Trust and Banking	5	5	1
automotive aftermarket	5	4	2
healthcare	5	5	3
Maufacturing, and more specifically, printing	5	5	2
Healthcare Medical Center	5	4	3
Retail/direct marketer	5	4	1
Air Compressor Manufacturing for Consumer & Commercial users	5	3	1
Rubber & Plastics	5	4	2
Retail/Supermarkets	5	5	1
Consulting	5	5	2
investment banking	5	4	3
Metal Stamping	5	4	2
Retail	5	5	2
OEM of plastic converting machinery	5	5	2
Retail/Industrial	5	4	2
Food Manufacturing	5	5	2
Construction - Industrial and Non-residential	5	2	2
Retail	5	5	2
Pharmaceutical	5	4	2
Agricultural Equipement Industry	5	5	4
Building Materials	2	3	4
Industrial and Aerospace Manufacturing	4	4	2
SUM	204	184	103
MIN	2	2	1
MAX	5	5	4
MEAN	4.4	4.0	2.2
MODE	5	4	2
MEDIAN	5	4	2
STANDARD DEVIATION	0.7	0.9	0.8
VARIANCE (VARP)	-	-	-
COVARIANCE OF VARIATION	0.2	0.2	0.4
CORRELATION COEEFICIENT (vs. OR)	-	-	-
SKEWNESS	-1.3	-0.8	0.8
KURTOSIS	2.1	0.4	0.4
Z-SCORE OF MIN - ROW # 50			
Z-SCORE OF MAX - ROW # 51			



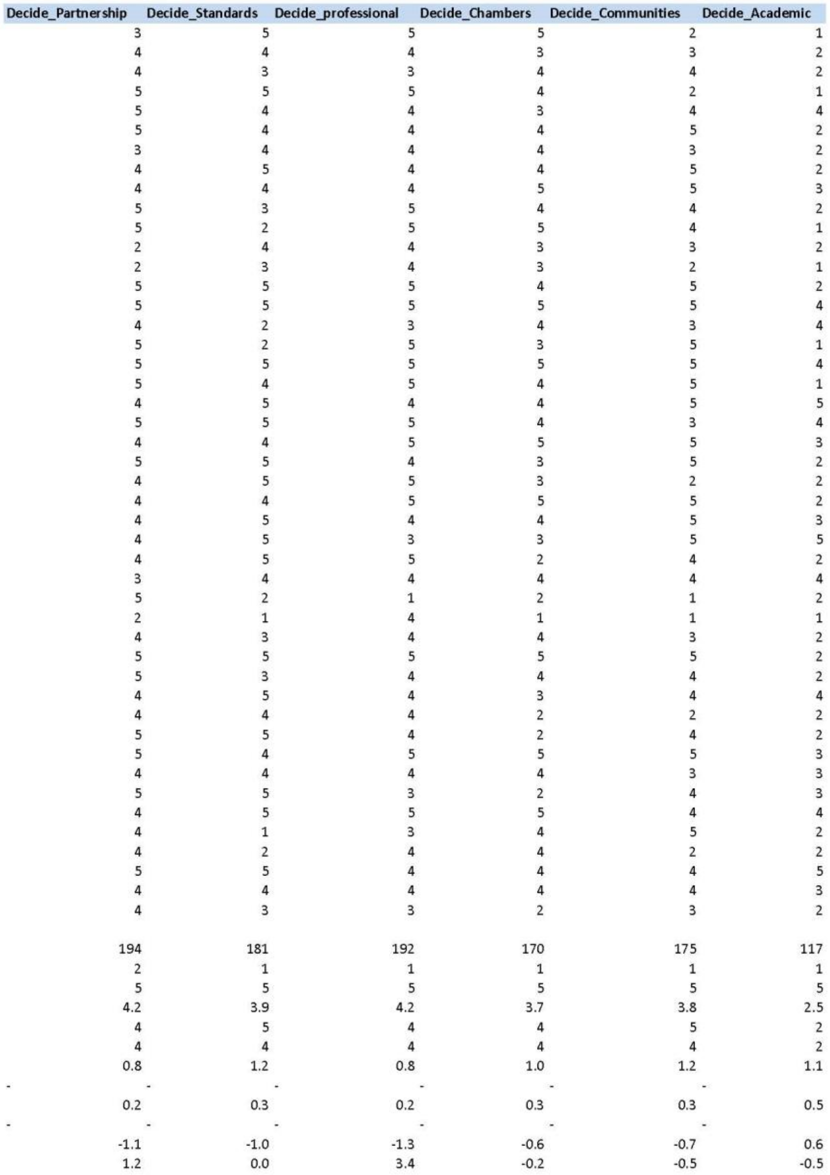
Create_Employees	Create_Facilities	Create_Suggest	Create_Experiment	Create_Insight	CreatePoints	CreatePossiblePoints
4	4	3	3	4	29	40
5	4	5	5	3	33	40
2	4	2	3	5	23	40
5	4	2	1	5	26	40
4	5	3	3	5	31	40
4	4	4	4	5	29	40
4	5	3	4	4	30	40
5	3	3	4	5	30	40
3	4	2	2	4	26	40
5	5	2	4	5	29	40
4	5	2	4	4	31	40
4	4	2	4	4	28	40
3	4	2	5	4	28	40
4	4	4	4	4	30	40
3	4	2	2	5	26	40
4	4	2	3	4	28	40
2	4	3	2	4	23	40
5	4	4	3	5	32	40
1	2	1	2	5	22	40
4	4	4	4	4	29	40
3	5	3	3	4	30	40
3	4	4	4	4	29	40
5	5	4	4	5	37	40
2	4	2	2	5	25	40
5	5	4	4	5	34	40
5	4	2	4	5	31	40
4	4	3	3	5	32	40
5	5	5	4	5	36	40
3	3	4	4	4	30	40
2	4	1	2	4	23	40
2	2	1	1	4	19	40
5	4	2	4	4	30	40
5	5	2	5	5	33	40
5	5	5	5	5	37	40
4	2	2	4	5	29	40
2	2	2	2	4	23	40
4	5	2	4	5	32	40
4	5	2	4	5	32	40
4	4	3	4	3	29	40
4	5	4	4	5	34	40
5	4	2	2	4	26	40
5	5	4	4	5	35	40
5	4	4	2	3	29	40
5	4	2	4	5	34	40
5	4	3	4	5	30	40
3	4	2	5	4	28	40
179	188	129	158	205		
1	2	1	1	3		
5	5	5	5	5		
3.9	4.1	2.8	3.4	4.5		
5	4	2	4	5		
4	4	2.5	4	5		
1.1	0.8	1.1	1.1	0.6		
-	-	-	-	-		
0.3	0.2	0.4	0.3	0.1		
-	-	-	-	-		
-0.8	-1.1	5.0	-0.6	-0.7		
-0.4	1.3	-0.8	-0.5	-0.4		



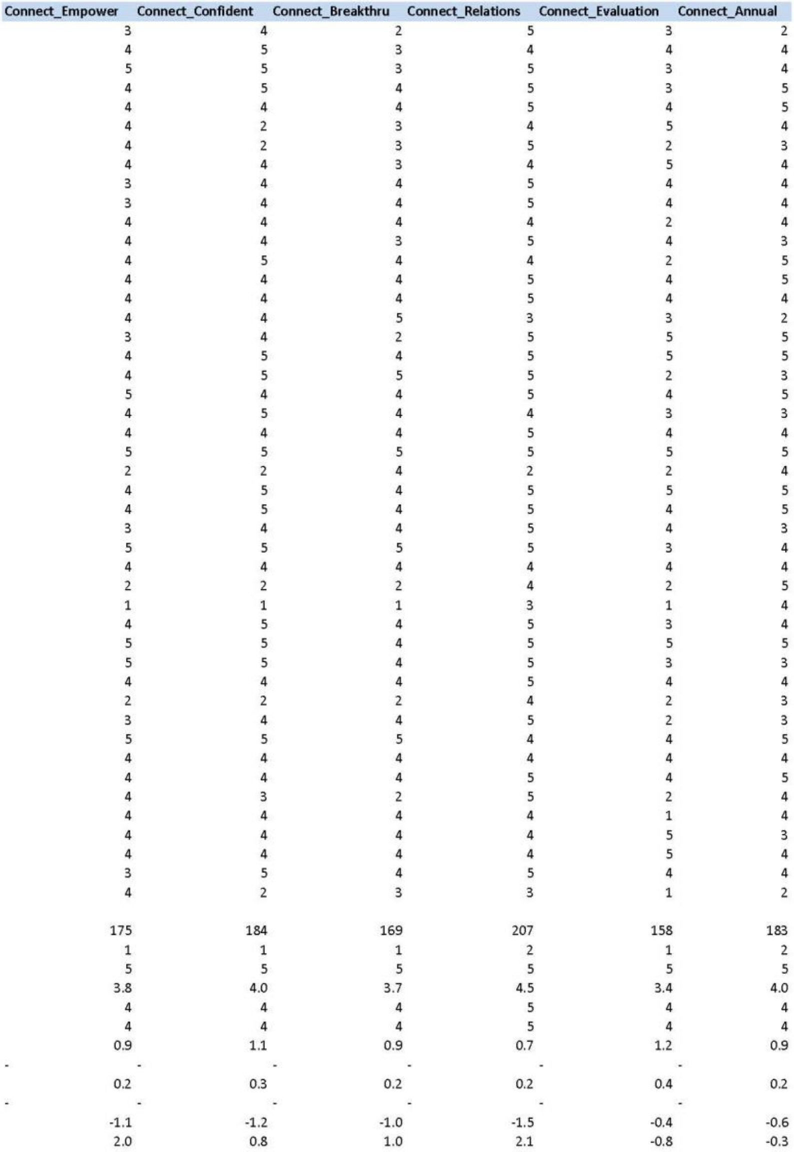
Exploit_Reflect	ExploitPoints	ExploitPossiblePoints	ExploitRatio	ExploitInteger	Exploit Z-SCORE	Decide_Alliances	Decide_CoOp
5	19	25	0.760	76		3	4
4	18	25	0.720	72		4	4
4	19	25	0.760	76		3	1
2	14	25	0.560	56		5	4
3	17	25	0.680	68		5	5
5	20	25	0.800	80		3	3
3	18	25	0.720	72		3	3
4	23	25	0.920	92		5	4
5	20	25	0.800	80		5	3
4	18	25	0.720	72		5	4
4	21	25	0.840	84		0	4
3	17	25	0.680	68		4	4
3	14	25	0.560	56		3	4
4	19	25	0.760	76		3	2
4	20	25	0.800	80		5	5
3	15	25	0.600	60		4	4
2	16	25	0.640	64		5	5
5	24	25	0.960	96		5	5
4	17	25	0.680	68		4	4
4	20	25	0.800	80		3	1
4	19	25	0.760	76		4	2
4	21	25	0.840	84		4	3
4	20	25	0.800	80		5	5
2	18	25	0.720	72		4	4
4	19	25	0.760	76		4	5
5	23	25	0.920	92		4	4
4	23	25	0.920	92		4	3
5	23	25	0.920	92		4	4
4	20	25	0.800	80		5	4
3	13	25	0.520	52		1	5
2	18	25	0.720	72		1	1
4	19	25	0.760	76		2	2
5	25	25	1.000	100		4	5
2	16	25	0.640	64		5	5
3	14	25	0.560	56		3	3
2	16	25	0.640	64		2	2
4	21	25	0.840	84		4	4
5	24	25	0.960	96		5	5
3	17	25	0.680	68		4	4
4	21	25	0.840	84		5	5
2	21	25	0.840	84		5	2
5	24	25	0.960	96		1	1
4	22	25	0.880	88		2	4
4	21	25	0.840	84		5	5
3	19	25	0.760	76		2	4
4	23	25	0.920	92		3	4
170			35.560	3556		169	168
2			0.520	52		0	1
5			1.000	100		5	5
3.7			0.773	77.3		3.7	3.7
4			0.760	76		5	4
4			0.760	76		4	4
1.0			0.119	11.9		1.3	1.2
-			-	139.2		-	-
0.3			0.154	0.2		0.4	0.3
-			0.247			-	-
-0.4			-0.152	-0.2		-0.9	-0.9
-0.7			-0.531	-0.5		0.3	-0.2
			-2.12				
			1.90				



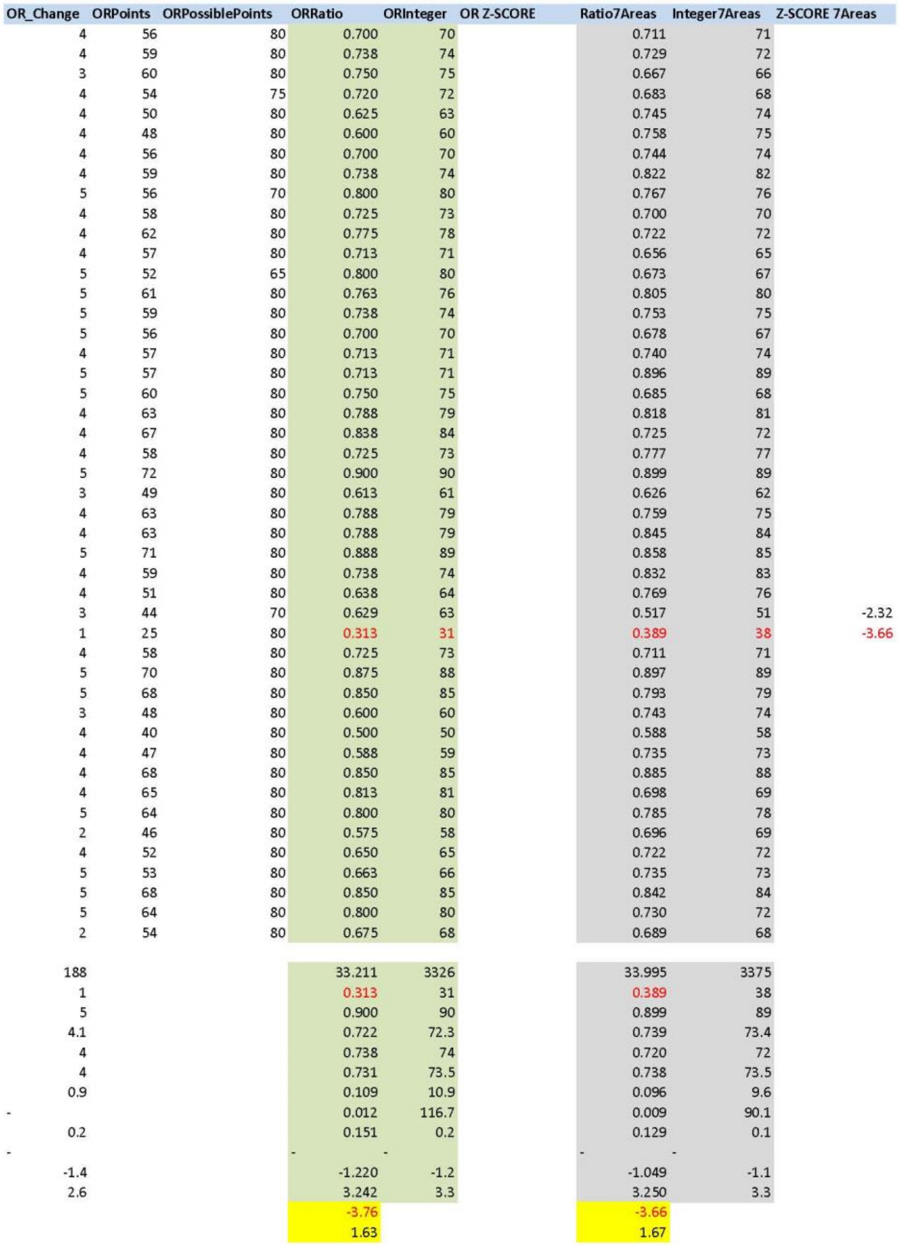
CreateRatio	CreateInteger	Create Z-SCORE	Exploit_References	Exploit_Simulate	Exploit_Consult	Exploit_ElectronicDB
0.725	73		5	2	2	5
0.825	83		4	3	3	4
0.575	58		4	2	5	4
0.650	65		4	1	3	4
0.775	78		4	3	4	3
0.725	73		4	2	5	4
0.750	75		4	3	4	4
0.750	75		5	4	5	5
0.650	65		4	3	4	4
0.725	73		2	4	4	4
0.775	78		4	4	5	4
0.700	70		5	3	4	2
0.700	70		1	4	4	2
0.750	75		4	3	4	4
0.650	65		4	4	5	3
0.700	70		4	2	4	2
0.575	58		5	2	2	5
0.800	80		4	5	5	5
0.550	55		5	1	4	3
0.725	73		4	3	5	4
0.750	75		4	4	3	4
0.725	73		4	3	5	5
0.925	93		4	4	4	4
0.625	63		5	4	5	2
0.850	85		5	4	4	2
0.775	78		5	4	5	4
0.800	80		5	5	5	4
0.900	90		4	4	5	5
0.750	75		4	4	4	4
0.575	58		4	2	3	1
0.475	48		4	5	3	4
0.750	75		3	4	4	4
0.825	83		5	5	5	5
0.925	93		5	2	5	2
0.725	73		3	2	4	2
0.575	58		4	4	4	2
0.800	80		5	4	4	4
0.800	80		5	4	5	5
0.725	73		3	4	4	3
0.850	85		5	4	4	4
0.650	65		5	5	5	4
0.875	88		5	4	5	5
0.725	73		4	4	5	5
0.850	85		4	3	5	5
0.750	75		4	4	4	4
0.700	70		4	5	5	5
33.750	3386		192	159	195	173
0.475	48		1	1	2	1
0.925	93		5	5	5	5
0.734	73.6		4.2	3.5	4.2	3.8
0.725	73		4	4	5	4
0.738	74		4	4	4	4
0.100	10.0		0.8	1.1	0.8	1.1
0.010	97.2		-	-	-	-
0.136	0.1		0.2	0.3	0.2	0.3
0.566	-		-	-	-	-
-0.331	-0.3		-1.6	-0.6	-1.0	-0.8
0.133	0.1		4.3	-0.4	0.6	-0.3
-2.59						
1.92						



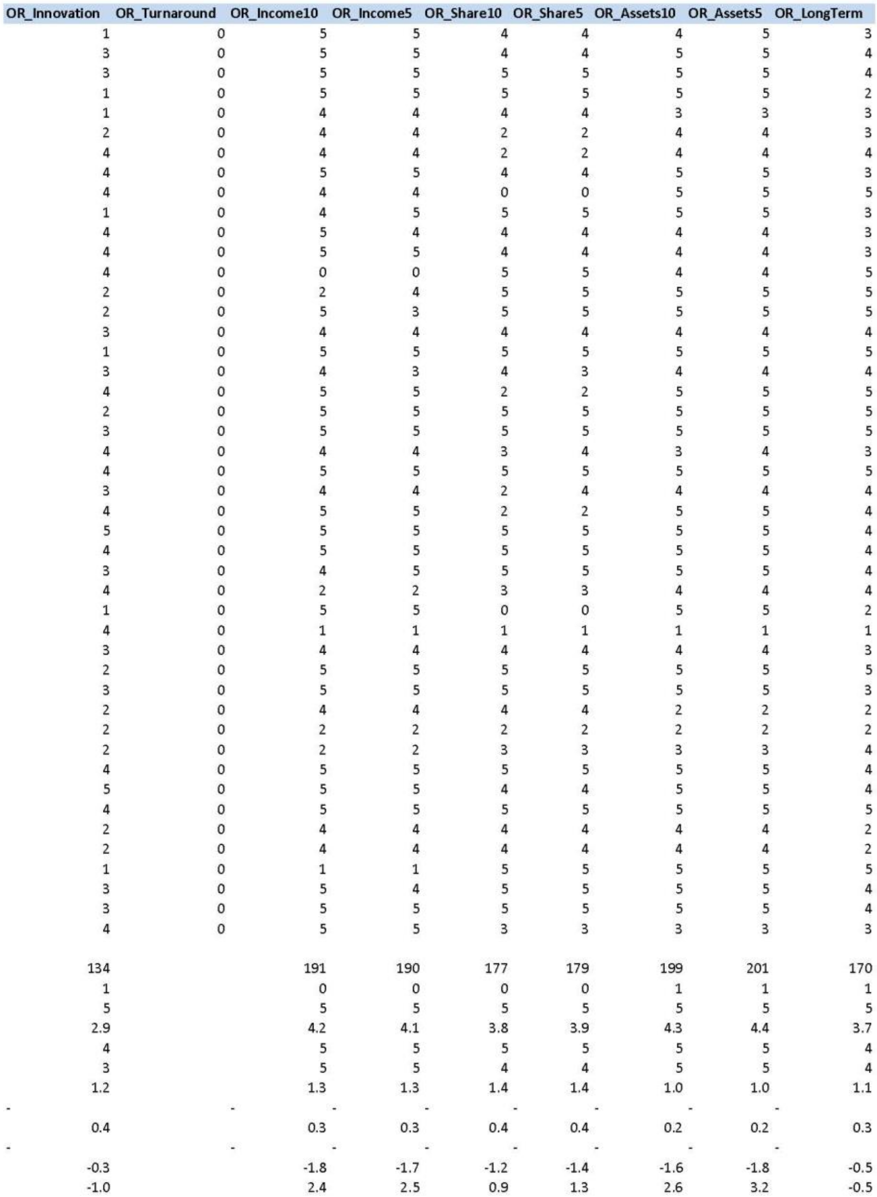
Decide_Partnership	Decide_Standards	Decide_professional	Decide_Chambers	Decide_Communities	Decide_Academic
3	5	5	5	2	1
4	4	4	3	3	2
4	3	3	4	4	2
5	5	5	4	2	1
5	4	4	3	4	4
5	4	4	4	5	2
3	4	4	4	3	2
4	5	4	4	5	2
4	4	4	5	5	3
5	3	5	4	4	2
5	2	5	5	4	1
2	4	4	3	3	2
2	3	4	3	2	1
5	5	5	4	5	2
5	5	5	5	5	4
4	2	3	4	3	4
5	2	5	3	5	1
5	5	5	5	5	4
5	4	5	4	5	1
4	5	4	4	5	5
5	5	5	4	3	4
4	4	5	5	5	3
5	5	4	3	5	2
4	5	5	3	2	2
4	4	5	5	5	2
4	5	4	4	5	3
4	5	3	3	5	5
4	5	5	2	4	2
3	4	4	4	4	4
5	2	1	2	1	2
2	1	4	1	1	1
4	3	4	4	3	2
5	5	5	5	5	2
5	3	4	4	4	2
4	5	4	3	4	4
4	4	4	2	2	2
5	5	4	2	4	2
5	4	5	5	5	3
4	4	4	4	3	3
5	5	3	2	4	3
4	5	5	5	4	4
4	1	3	4	5	2
4	2	4	4	2	2
5	5	4	4	4	5
4	4	4	4	4	3
4	3	3	2	3	2
194	181	192	170	175	117
2	1	1	1	1	1
5	5	5	5	5	5
4.2	3.9	4.2	3.7	3.8	2.5
4	5	4	4	5	2
4	4	4	4	4	2
0.8	1.2	0.8	1.0	1.2	1.1
-	-	-	-	-	-
0.2	0.3	0.2	0.3	0.3	0.5
-	-	-	-	-	-
-1.1	-1.0	-1.3	-0.6	-0.7	0.6
1.2	0.0	3.4	-0.2	-0.5	-0.5



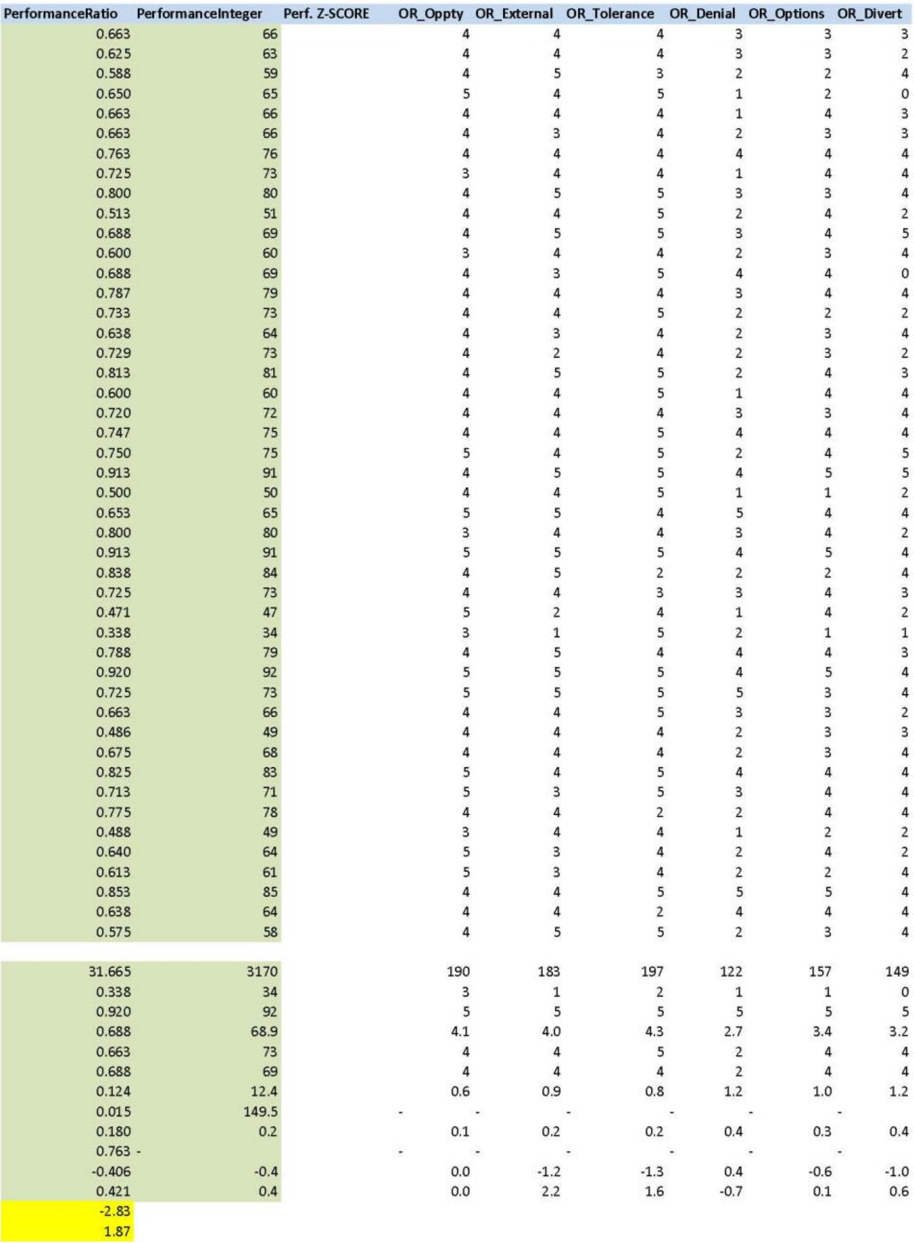
Connect_Empower	Connect_Confident	Connect_Breakthru	Connect_Relations	Connect_Evaluation	Connect_Annual
3	4	2	5	3	2
4	5	3	4	4	4
5	5	3	5	3	4
4	5	4	5	3	5
4	4	4	5	4	5
4	2	3	4	5	4
4	2	3	5	2	3
4	4	3	4	5	4
3	4	4	5	4	4
3	4	4	5	4	4
4	4	4	4	2	4
4	4	3	5	4	3
4	5	4	4	2	5
4	4	4	5	4	5
4	4	4	5	4	4
4	4	5	3	3	2
3	4	2	5	5	5
4	5	4	5	5	5
4	5	5	5	2	3
5	4	4	5	4	5
4	5	4	4	3	3
4	4	4	5	4	4
5	5	5	5	5	5
2	2	4	2	2	4
4	5	4	5	5	5
4	5	4	5	4	5
3	4	4	5	4	3
5	5	5	5	3	4
4	4	4	4	4	4
2	2	2	4	2	5
1	1	1	3	1	4
4	5	4	5	3	4
5	5	4	5	5	5
5	5	4	5	3	3
4	4	4	5	4	4
2	2	2	4	2	3
3	4	4	5	2	3
5	5	5	4	4	5
4	4	4	4	4	4
4	4	4	5	4	5
4	3	2	5	2	4
4	4	4	4	1	4
4	4	4	4	5	3
4	4	4	4	5	4
3	5	4	5	4	4
4	2	3	3	1	2
175	184	169	207	158	183
1	1	1	2	1	2
5	5	5	5	5	5
3.8	4.0	3.7	4.5	3.4	4.0
4	4	4	5	4	4
4	4	4	5	4	4
0.9	1.1	0.9	0.7	1.2	0.9
-	-	-	-	-	-
0.2	0.3	0.2	0.2	0.4	0.2
-	-	-	-	-	-
-1.1	-1.2	-1.0	-1.5	-0.4	-0.6
2.0	0.8	1.0	2.1	-0.8	-0.3



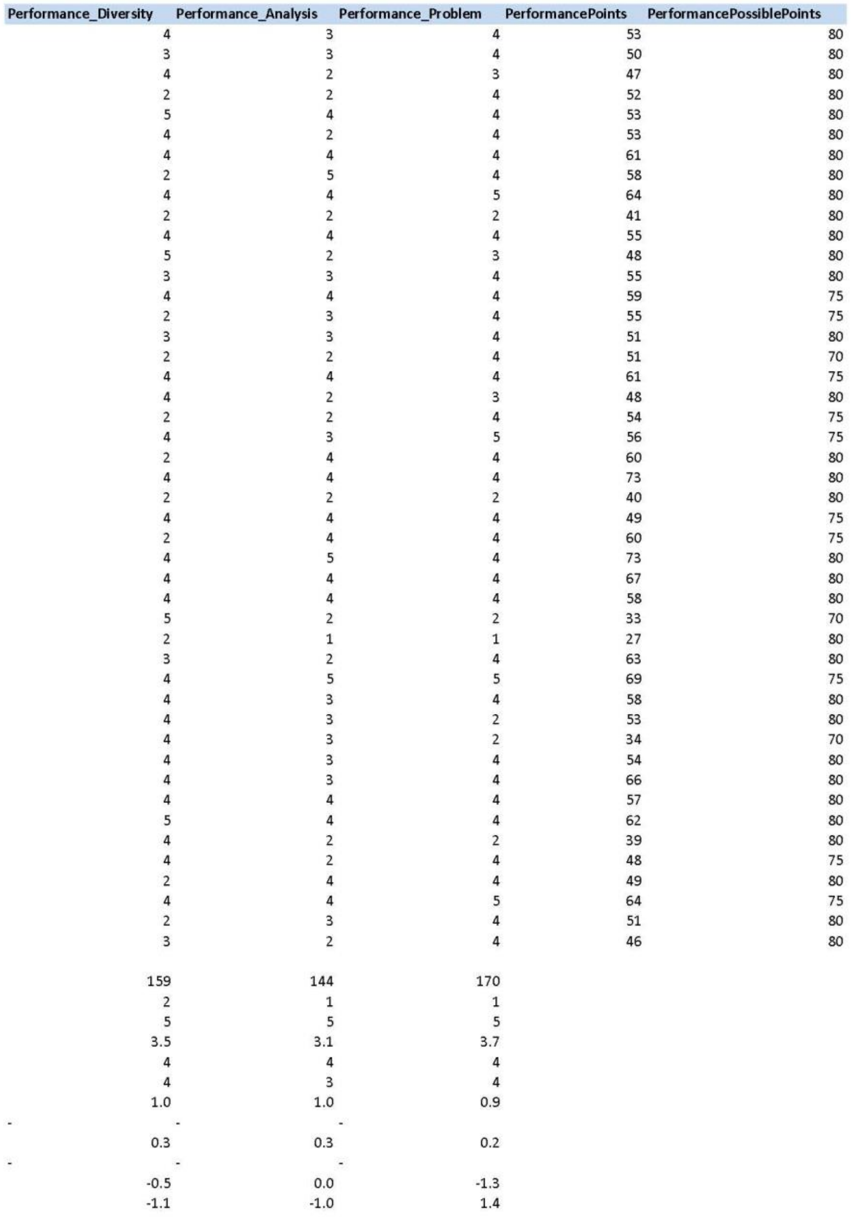
OR_Change	ORPoints	ORPossiblePoints	ORRatio	ORInteger	OR Z-SCORE	Ratio7Areas	Integer7Areas	Z-SCORE 7Areas
4	56	80	0.700	70		0.711	71	
4	59	80	0.738	74		0.729	72	
3	60	80	0.750	75		0.667	66	
4	54	75	0.720	72		0.683	68	
4	50	80	0.625	63		0.745	74	
4	48	80	0.600	60		0.758	75	
4	56	80	0.700	70		0.744	74	
4	59	80	0.738	74		0.822	82	
5	56	70	0.800	80		0.767	76	
4	58	80	0.725	73		0.700	70	
4	62	80	0.775	78		0.722	72	
4	57	80	0.713	71		0.656	65	
5	52	65	0.800	80		0.673	67	
5	61	80	0.763	76		0.805	80	
5	59	80	0.738	74		0.753	75	
5	56	80	0.700	70		0.678	67	
4	57	80	0.713	71		0.740	74	
5	57	80	0.713	71		0.896	89	
5	60	80	0.750	75		0.685	68	
4	63	80	0.788	79		0.818	81	
4	67	80	0.838	84		0.725	72	
4	58	80	0.725	73		0.777	77	
5	72	80	0.900	90		0.899	89	
3	49	80	0.613	61		0.626	62	
4	63	80	0.788	79		0.759	75	
4	63	80	0.788	79		0.845	84	
5	71	80	0.888	89		0.858	85	
4	59	80	0.738	74		0.832	83	
4	51	80	0.638	64		0.769	76	
3	44	70	0.629	63		0.517	51	-2.32
1	25	80	0.313	31		0.389	38	-3.66
4	58	80	0.725	73		0.711	71	
5	70	80	0.875	88		0.897	89	
5	68	80	0.850	85		0.793	79	
3	48	80	0.600	60		0.743	74	
4	40	80	0.500	50		0.588	58	
4	47	80	0.588	59		0.735	73	
4	68	80	0.850	85		0.885	88	
4	65	80	0.813	81		0.698	69	
5	64	80	0.800	80		0.785	78	
2	46	80	0.575	58		0.696	69	
4	52	80	0.650	65		0.722	72	
5	53	80	0.663	66		0.735	73	
5	68	80	0.850	85		0.842	84	
5	64	80	0.800	80		0.730	72	
2	54	80	0.675	68		0.689	68	
188			33.211	3326		33.995	3375	
1			0.313	31		0.389	38	
5			0.900	90		0.899	89	
4.1			0.722	72.3		0.739	73.4	
4			0.738	74		0.720	72	
4			0.731	73.5		0.738	73.5	
0.9			0.109	10.9		0.096	9.6	
-			0.012	116.7		0.009	90.1	
0.2			0.151	0.2		0.129	0.1	
-			-	-		-	-	
-1.4			-1.220	-1.2		-1.049	-1.1	
2.6			3.242	3.3		3.250	3.3	
			-3.76			-3.66		
			1.63			1.67		



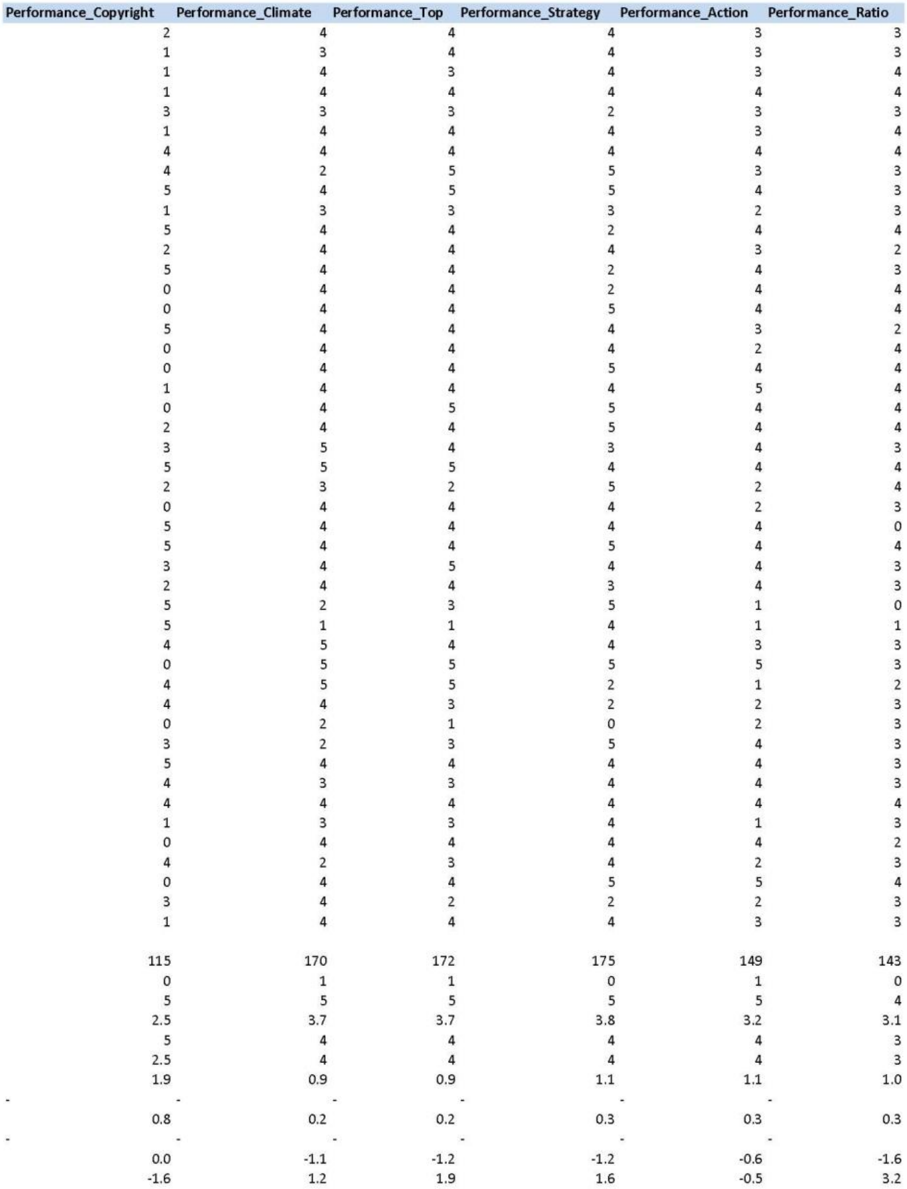
OR_Innovation	OR_Turnaround	OR_Income10	OR_Income5	OR_Share10	OR_Share5	OR_Assets10	OR_Assets5	OR_LongTerm
1	0	5	5	4	4	4	5	3
3	0	5	5	4	4	5	5	4
3	0	5	5	5	5	5	5	4
1	0	5	5	5	5	5	5	2
1	0	4	4	4	4	3	3	3
2	0	4	4	2	2	4	4	3
4	0	4	4	2	2	4	4	4
4	0	5	5	4	4	5	5	3
4	0	4	4	0	0	5	5	5
1	0	4	5	5	5	5	5	3
4	0	5	4	4	4	4	4	3
4	0	5	5	4	4	4	4	3
4	0	0	0	5	5	4	4	5
2	0	2	4	5	5	5	5	5
2	0	5	3	5	5	5	5	5
3	0	4	4	4	4	4	4	4
1	0	5	5	5	5	5	5	5
3	0	4	3	4	3	4	4	4
4	0	5	5	2	2	5	5	5
2	0	5	5	5	5	5	5	5
3	0	5	5	5	5	5	5	5
4	0	4	4	3	4	3	4	3
4	0	5	5	5	5	5	5	5
3	0	4	4	2	4	4	4	4
4	0	5	5	2	2	5	5	4
5	0	5	5	5	5	5	5	4
4	0	5	5	5	5	5	5	4
3	0	4	5	5	5	5	5	4
4	0	2	2	3	3	4	4	4
1	0	5	5	0	0	5	5	2
4	0	1	1	1	1	1	1	1
3	0	4	4	4	4	4	4	3
2	0	5	5	5	5	5	5	5
3	0	5	5	5	5	5	5	3
2	0	4	4	4	4	2	2	2
2	0	2	2	2	2	2	2	2
2	0	2	2	3	3	3	3	4
4	0	5	5	5	5	5	5	4
5	0	5	5	4	4	5	5	4
4	0	5	5	5	5	5	5	5
2	0	4	4	4	4	4	4	2
2	0	4	4	4	4	4	4	2
1	0	1	1	5	5	5	5	5
3	0	5	4	5	5	5	5	4
3	0	5	5	5	5	5	5	4
4	0	5	5	3	3	3	3	3
134		191	190	177	179	199	201	170
1		0	0	0	0	1	1	1
5		5	5	5	5	5	5	5
2.9		4.2	4.1	3.8	3.9	4.3	4.4	3.7
4		5	5	5	5	5	5	4
3		5	5	4	4	5	5	4
1.2		1.3	1.3	1.4	1.4	1.0	1.0	1.1
-	-	-	-	-	-	-	-	
0.4		0.3	0.3	0.4	0.4	0.2	0.2	0.3
-	-	-	-	-	-	-	-	
-0.3		-1.8	-1.7	-1.2	-1.4	-1.6	-1.8	-0.5
-1.0		2.4	2.5	0.9	1.3	2.6	3.2	-0.5



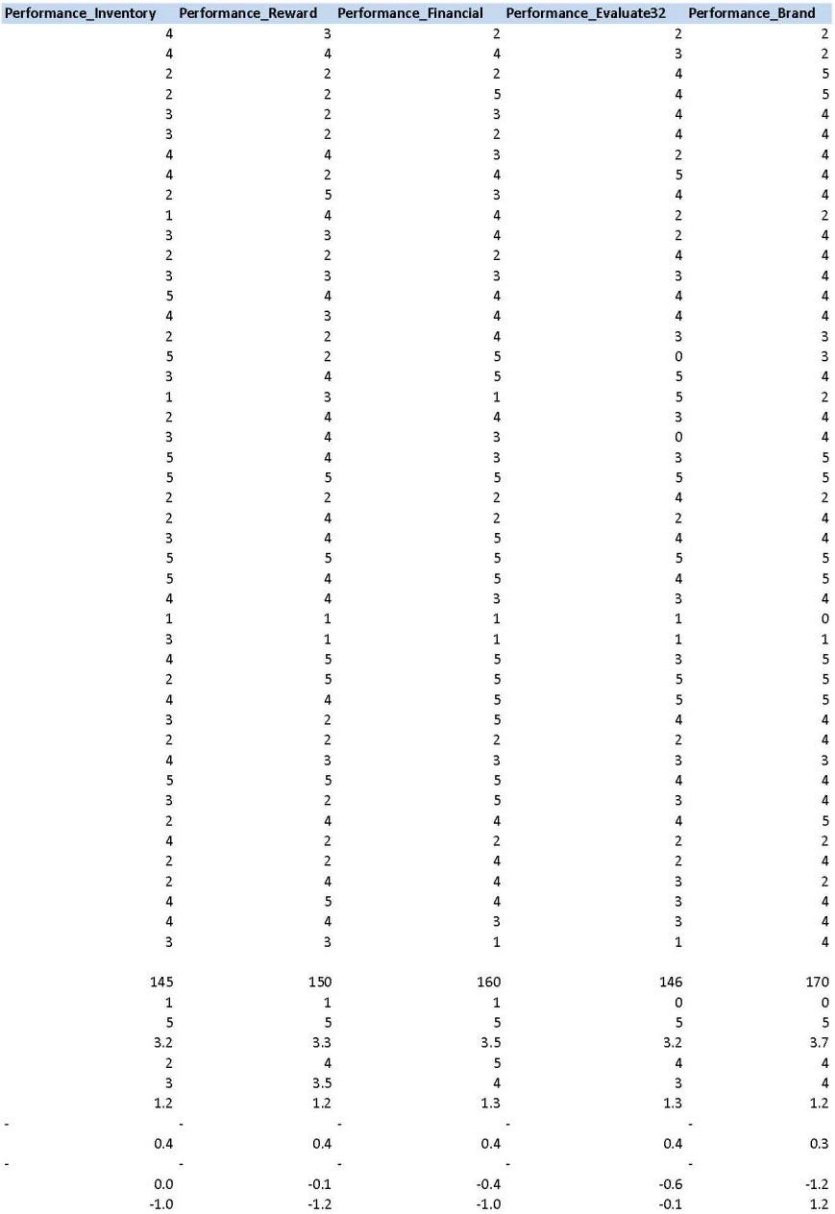
PerformanceRatio	PerformanceInteger	Perf. Z-SCORE	OR_Oppty	OR_External	OR_Tolerance	OR_Denial	OR_Options	OR_Divert
0.663	66		4	4	4	3	3	3
0.625	63		4	4	4	3	3	2
0.588	59		4	5	3	2	2	4
0.650	65		5	4	5	1	2	0
0.663	66		4	4	4	1	4	3
0.663	66		4	3	4	2	3	3
0.763	76		4	4	4	4	4	4
0.725	73		3	4	4	1	4	4
0.800	80		4	5	5	3	3	4
0.513	51		4	4	5	2	4	2
0.688	69		4	5	5	3	4	5
0.600	60		3	4	4	2	3	4
0.688	69		4	3	5	4	4	0
0.787	79		4	4	4	3	4	4
0.733	73		4	4	5	2	2	2
0.638	64		4	3	4	2	3	4
0.729	73		4	2	4	2	3	2
0.813	81		4	5	5	2	4	3
0.600	60		4	4	5	1	4	4
0.720	72		4	4	4	3	3	4
0.747	75		4	4	5	4	4	4
0.750	75		5	4	5	2	4	5
0.913	91		4	5	5	4	5	5
0.500	50		4	4	5	1	1	2
0.653	65		5	5	4	5	4	4
0.800	80		3	4	4	3	4	2
0.913	91		5	5	5	4	5	4
0.838	84		4	5	2	2	2	4
0.725	73		4	4	3	3	4	3
0.471	47		5	2	4	1	4	2
0.338	34		3	1	5	2	1	1
0.788	79		4	5	4	4	4	3
0.920	92		5	5	5	4	5	4
0.725	73		5	5	5	5	3	4
0.663	66		4	4	5	3	3	2
0.486	49		4	4	4	2	3	3
0.675	68		4	4	4	2	3	4
0.825	83		5	4	5	4	4	4
0.713	71		5	3	5	3	4	4
0.775	78		4	4	2	2	4	4
0.488	49		3	4	4	1	2	2
0.640	64		5	3	4	2	4	2
0.613	61		5	3	4	2	2	4
0.853	85		4	4	5	5	5	4
0.638	64		4	4	2	4	4	4
0.575	58		4	5	5	2	3	4
31.665	3170		190	183	197	122	157	149
0.338	34		3	1	2	1	1	0
0.920	92		5	5	5	5	5	5
0.688	68.9		4.1	4.0	4.3	2.7	3.4	3.2
0.663	73		4	4	5	2	4	4
0.688	69		4	4	4	2	4	4
0.124	12.4		0.6	0.9	0.8	1.2	1.0	1.2
0.015	149.5		-	-	-	-	-	-
0.180	0.2		0.1	0.2	0.2	0.4	0.3	0.4
0.763	-		-	-	-	-	-	-
-0.406	-0.4		0.0	-1.2	-1.3	0.4	-0.6	-1.0
0.421	0.4		0.0	2.2	1.6	-0.7	0.1	0.6
-2.83								
1.87								



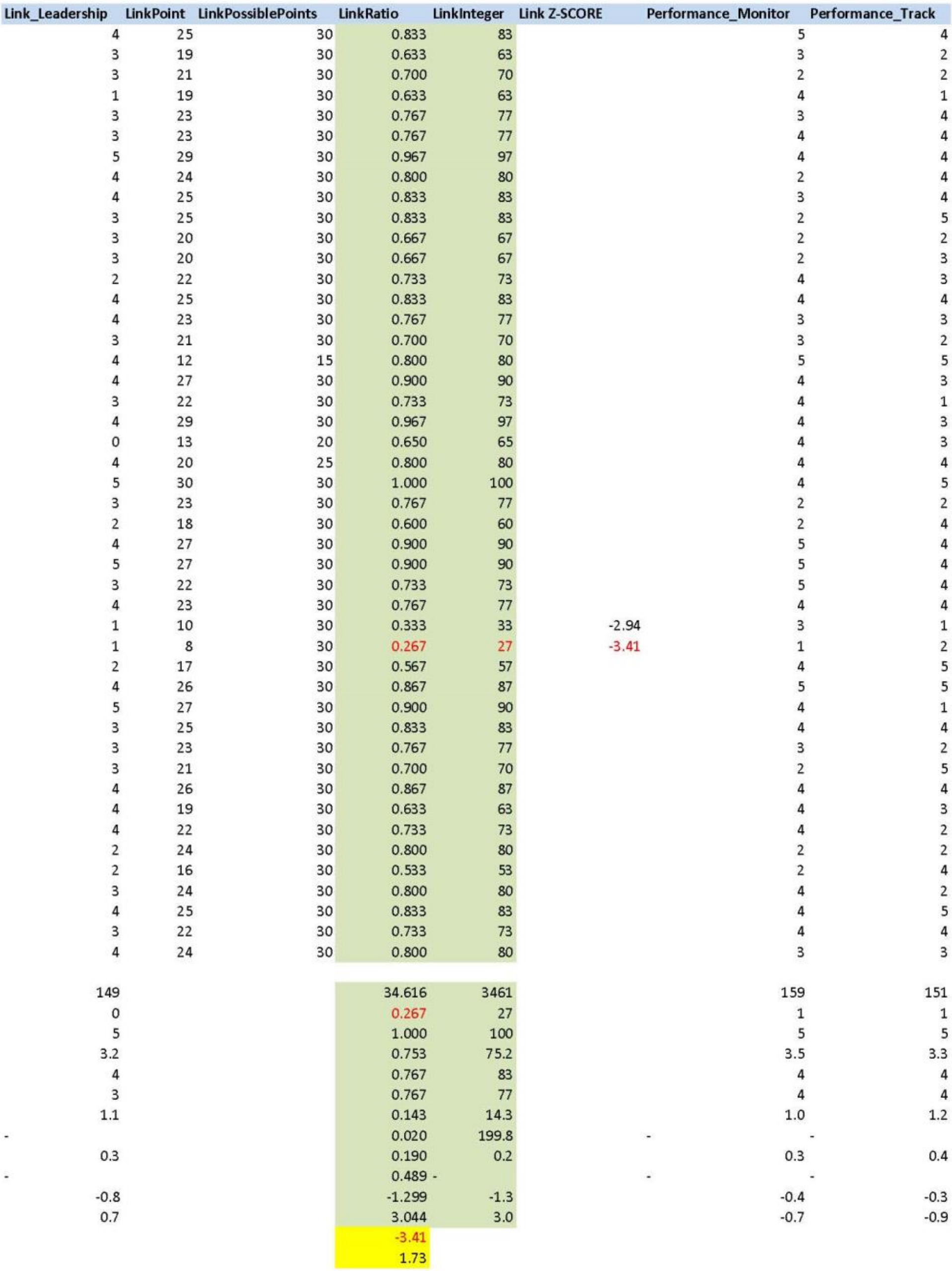
Performance_Diversity	Performance_Analysis	Performance_Problem	PerformancePoints	PerformancePossiblePoints
4	3	4	53	80
3	3	4	50	80
4	2	3	47	80
2	2	4	52	80
5	4	4	53	80
4	2	4	53	80
4	4	4	61	80
2	5	4	58	80
4	4	5	64	80
2	2	2	41	80
4	4	4	55	80
5	2	3	48	80
3	3	4	55	80
4	4	4	59	75
2	3	4	55	75
3	3	4	51	80
2	2	4	51	70
4	4	4	61	75
4	2	3	48	80
2	2	4	54	75
4	3	5	56	75
2	4	4	60	80
4	4	4	73	80
2	2	2	40	80
4	4	4	49	75
2	4	4	60	75
4	5	4	73	80
4	4	4	67	80
4	4	4	58	80
5	2	2	33	70
2	1	1	27	80
3	2	4	63	80
4	5	5	69	75
4	3	4	58	80
4	3	2	53	80
4	3	2	34	70
4	3	4	54	80
4	3	4	66	80
4	4	4	57	80
5	4	4	62	80
4	2	2	39	80
4	2	4	48	75
2	4	4	49	80
4	4	5	64	75
2	3	4	51	80
3	2	4	46	80
159	144	170		
2	1	1		
5	5	5		
3.5	3.1	3.7		
4	4	4		
4	3	4		
1.0	1.0	0.9		
-	-	-		
0.3	0.3	0.2		
-	-	-		
-0.5	0.0	-1.3		
-1.1	-1.0	1.4		



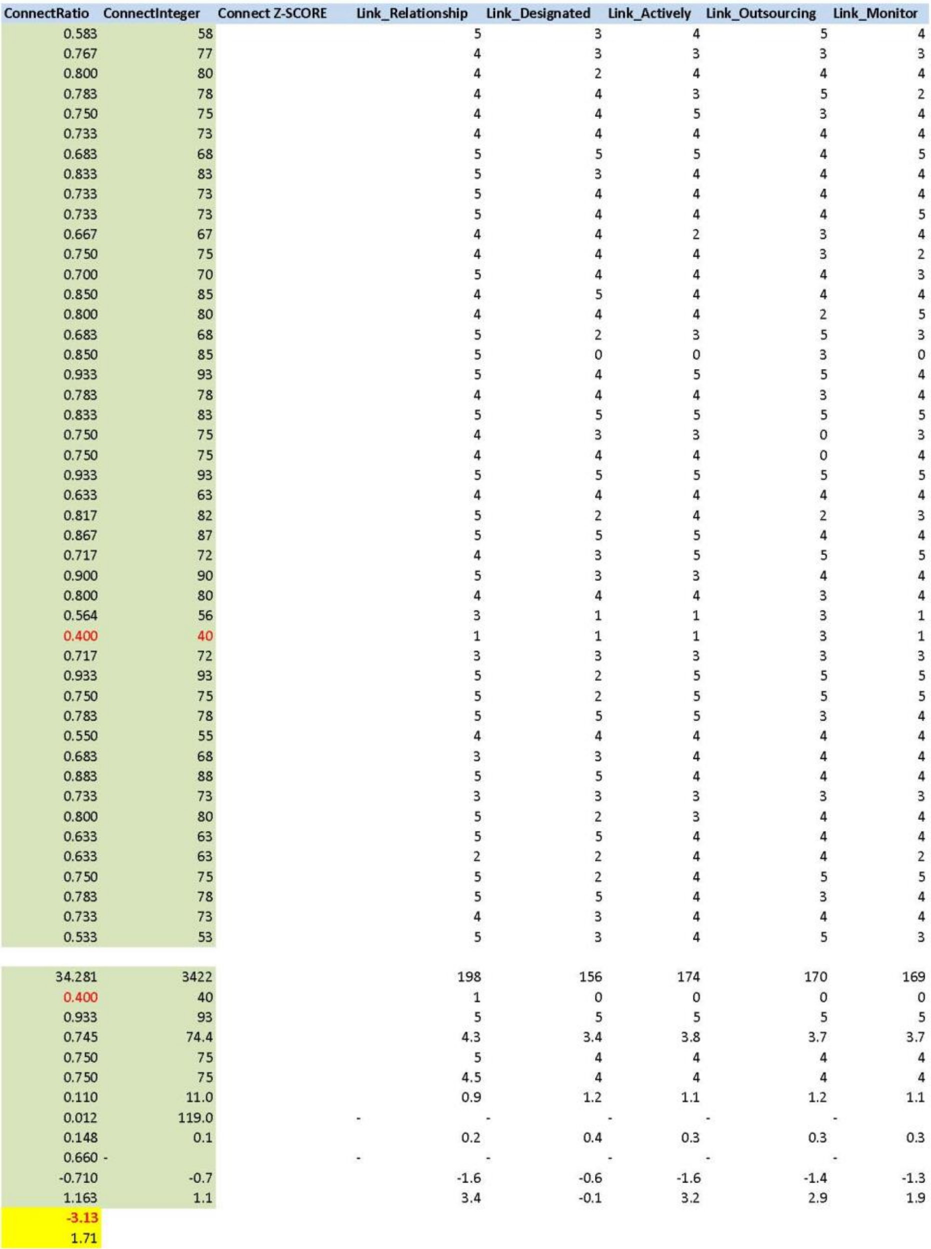
Performance_Copyright	Performance_Climate	Performance_Top	Performance_Strategy	Performance_Action	Performance_Ratio
2	4	4	4	3	3
1	3	4	4	3	3
1	4	3	4	3	4
1	4	4	4	4	4
3	3	3	2	3	3
1	4	4	4	3	4
4	4	4	4	4	4
4	2	5	5	3	3
5	4	5	5	4	3
1	3	3	3	2	3
5	4	4	2	4	4
2	4	4	4	3	2
5	4	4	2	4	3
0	4	4	2	4	4
0	4	4	5	4	4
5	4	4	4	3	2
0	4	4	4	2	4
0	4	4	5	4	4
1	4	4	4	5	4
0	4	5	5	4	4
2	4	4	5	4	4
3	5	4	3	4	3
5	5	5	4	4	4
2	3	2	5	2	4
0	4	4	4	2	3
5	4	4	4	4	0
5	4	4	5	4	4
3	4	5	4	4	3
2	4	4	3	4	3
5	2	3	5	1	0
5	1	1	4	1	1
4	5	4	4	3	3
0	5	5	5	5	3
4	5	5	2	1	2
4	4	3	2	2	3
0	2	1	0	2	3
3	2	3	5	4	3
5	4	4	4	4	3
4	3	3	4	4	3
4	4	4	4	4	4
1	3	3	4	1	3
0	4	4	4	4	2
4	2	3	4	2	3
0	4	4	5	5	4
3	4	2	2	2	3
1	4	4	4	3	3
115	170	172	175	149	143
0	1	1	0	1	0
5	5	5	5	5	4
2.5	3.7	3.7	3.8	3.2	3.1
5	4	4	4	4	3
2.5	4	4	4	4	3
1.9	0.9	0.9	1.1	1.1	1.0
-	-	-	-	-	-
0.8	0.2	0.2	0.3	0.3	0.3
-	-	-	-	-	-
0.0	-1.1	-1.2	-1.2	-0.6	-1.6
-1.6	1.2	1.9	1.6	-0.5	3.2



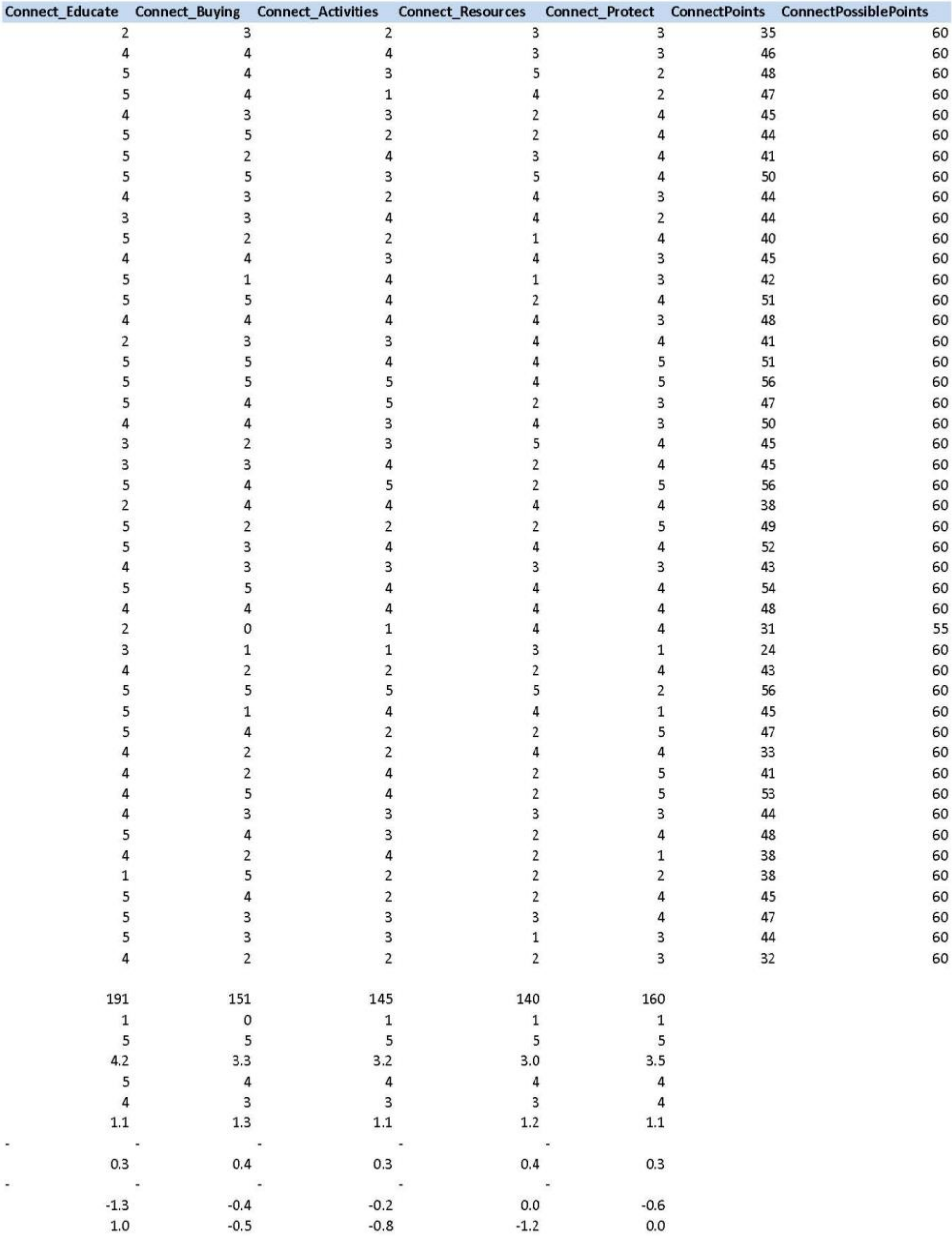
Performance_Inventory	Performance_Reward	Performance_Financial	Performance_Evaluate32	Performance_Brand
4	3	2	2	2
4	4	4	3	2
2	2	2	4	5
2	2	5	4	5
3	2	3	4	4
3	2	2	4	4
4	4	3	2	4
4	2	4	5	4
2	5	3	4	4
1	4	4	2	2
3	3	4	2	4
2	2	2	4	4
3	3	3	3	4
5	4	4	4	4
4	3	4	4	4
2	2	4	3	3
5	2	5	0	3
3	4	5	5	4
1	3	1	5	2
2	4	4	3	4
3	4	3	0	4
5	4	3	3	5
5	5	5	5	5
2	2	2	4	2
2	4	2	2	4
3	4	5	4	4
5	5	5	5	5
5	4	5	4	5
4	4	3	3	4
1	1	1	1	0
3	1	1	1	1
4	5	5	3	5
2	5	5	5	5
4	4	5	5	5
3	2	5	4	4
2	2	2	2	4
4	3	3	3	3
5	5	5	4	4
3	2	5	3	4
2	4	4	4	5
4	2	2	2	2
2	2	4	2	4
2	4	4	3	2
4	5	4	3	4
4	4	3	3	4
3	3	1	1	4
145	150	160	146	170
1	1	1	0	0
5	5	5	5	5
3.2	3.3	3.5	3.2	3.7
2	4	5	4	4
3	3.5	4	3	4
1.2	1.2	1.3	1.3	1.2
-	-	-	-	-
0.4	0.4	0.4	0.4	0.3
-	-	-	-	-
0.0	-0.1	-0.4	-0.6	-1.2
-1.0	-1.2	-1.0	-0.1	1.2



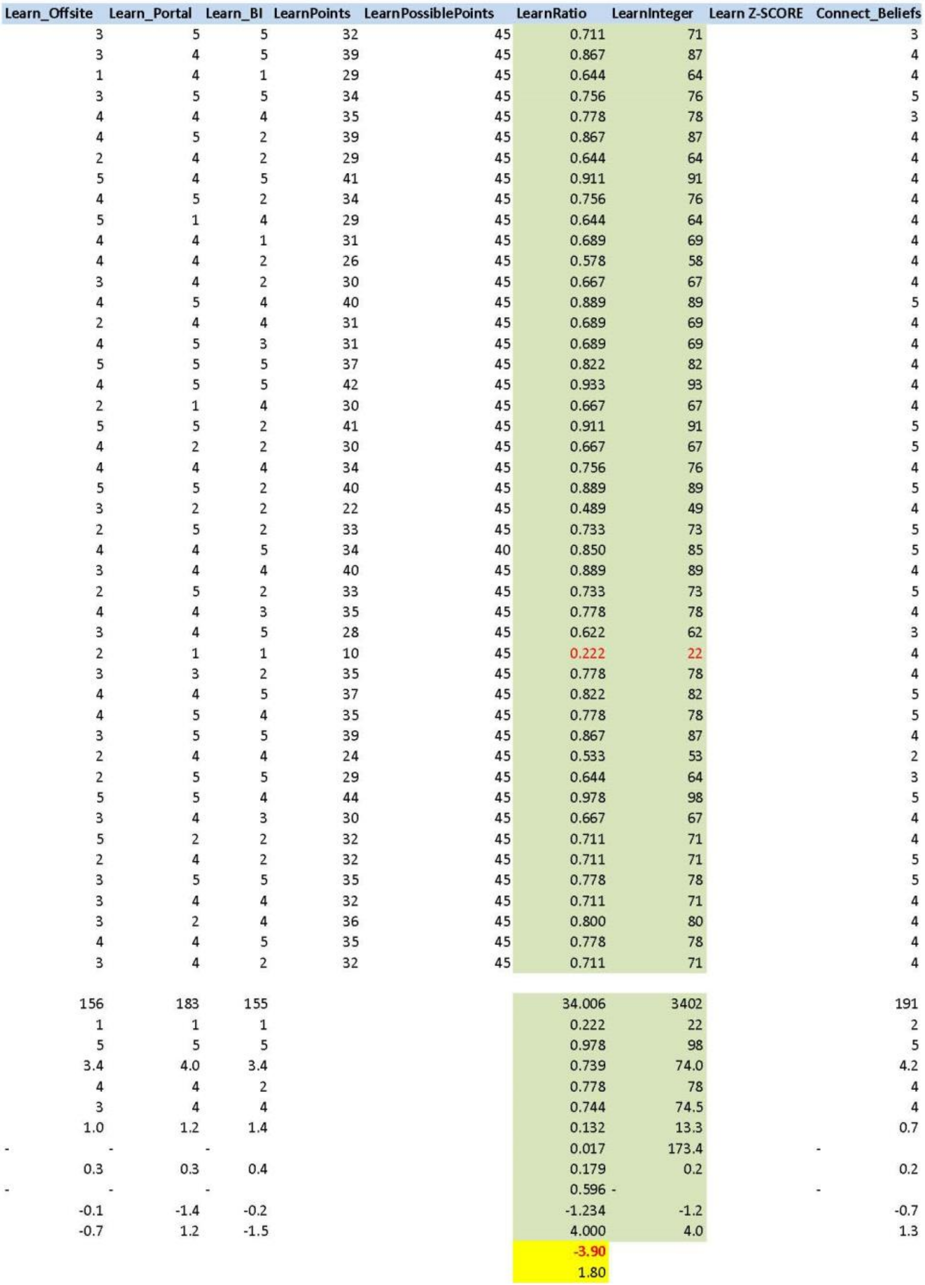
Link_Leadership	LinkPoint	LinkPossiblePoints	LinkRatio	LinkInteger	Link Z-SCORE	Performance_Monitor	Performance_Track
4	25	30	0.833	83		5	4
3	19	30	0.633	63		3	2
3	21	30	0.700	70		2	2
1	19	30	0.633	63		4	1
3	23	30	0.767	77		3	4
3	23	30	0.767	77		4	4
5	29	30	0.967	97		4	4
4	24	30	0.800	80		2	4
4	25	30	0.833	83		3	4
3	25	30	0.833	83		2	5
3	20	30	0.667	67		2	2
3	20	30	0.667	67		2	3
2	22	30	0.733	73		4	3
4	25	30	0.833	83		4	4
4	23	30	0.767	77		3	3
3	21	30	0.700	70		3	2
4	12	15	0.800	80		5	5
4	27	30	0.900	90		4	3
3	22	30	0.733	73		4	1
4	29	30	0.967	97		4	3
0	13	20	0.650	65		4	3
4	20	25	0.800	80		4	4
5	30	30	1.000	100		4	5
3	23	30	0.767	77		2	2
2	18	30	0.600	60		2	4
4	27	30	0.900	90		5	4
5	27	30	0.900	90		5	4
3	22	30	0.733	73		5	4
4	23	30	0.767	77		4	4
1	10	30	0.333	33	-2.94	3	1
1	8	30	0.267	27	-3.41	1	2
2	17	30	0.567	57		4	5
4	26	30	0.867	87		5	5
5	27	30	0.900	90		4	1
3	25	30	0.833	83		4	4
3	23	30	0.767	77		3	2
3	21	30	0.700	70		2	5
4	26	30	0.867	87		4	4
4	19	30	0.633	63		4	3
4	22	30	0.733	73		4	2
2	24	30	0.800	80		2	2
2	16	30	0.533	53		2	4
3	24	30	0.800	80		4	2
4	25	30	0.833	83		4	5
3	22	30	0.733	73		4	4
4	24	30	0.800	80		3	3
149			34.616	3461		159	151
0			0.267	27		1	1
5			1.000	100		5	5
3.2			0.753	75.2		3.5	3.3
4			0.767	83		4	4
3			0.767	77		4	4
1.1			0.143	14.3		1.0	1.2
-			0.020	199.8		-	-
0.3			0.190	0.2		0.3	0.4
-			0.489	-		-	-
-0.8			-1.299	-1.3		-0.4	-0.3
0.7			3.044	3.0		-0.7	-0.9
			-3.41				
			1.73				



ConnectRatio	ConnectInteger	Connect Z-SCORE	Link_Relationship	Link_Designated	Link_Actively	Link_Outsourcing	Link_Monitor
0.583	58		5	3	4	5	4
0.767	77		4	3	3	3	3
0.800	80		4	2	4	4	4
0.783	78		4	4	3	5	2
0.750	75		4	4	5	3	4
0.733	73		4	4	4	4	4
0.683	68		5	5	5	4	5
0.833	83		5	3	4	4	4
0.733	73		5	4	4	4	4
0.733	73		5	4	4	4	5
0.667	67		4	4	2	3	4
0.750	75		4	4	4	3	2
0.700	70		5	4	4	4	3
0.850	85		4	5	4	4	4
0.800	80		4	4	4	2	5
0.683	68		5	2	3	5	3
0.850	85		5	0	0	3	0
0.933	93		5	4	5	5	4
0.783	78		4	4	4	3	4
0.833	83		5	5	5	5	5
0.750	75		4	3	3	0	3
0.750	75		4	4	4	0	4
0.933	93		5	5	5	5	5
0.633	63		4	4	4	4	4
0.817	82		5	2	4	2	3
0.867	87		5	5	5	4	4
0.717	72		4	3	5	5	5
0.900	90		5	3	3	4	4
0.800	80		4	4	4	3	4
0.564	56		3	1	1	3	1
0.400	40		1	1	1	3	1
0.717	72		3	3	3	3	3
0.933	93		5	2	5	5	5
0.750	75		5	2	5	5	5
0.783	78		5	5	5	3	4
0.550	55		4	4	4	4	4
0.683	68		3	3	4	4	4
0.883	88		5	5	4	4	4
0.733	73		3	3	3	3	3
0.800	80		5	2	3	4	4
0.633	63		5	5	4	4	4
0.633	63		2	2	4	4	2
0.750	75		5	2	4	5	5
0.783	78		5	5	4	3	4
0.733	73		4	3	4	4	4
0.533	53		5	3	4	5	3
34.281	3422		198	156	174	170	169
0.400	40		1	0	0	0	0
0.933	93		5	5	5	5	5
0.745	74.4		4.3	3.4	3.8	3.7	3.7
0.750	75		5	4	4	4	4
0.750	75		4.5	4	4	4	4
0.110	11.0		0.9	1.2	1.1	1.2	1.1
0.012	119.0	-	-	-	-	-	
0.148	0.1		0.2	0.4	0.3	0.3	0.3
0.660	-	-	-	-	-	-	
-0.710	-0.7		-1.6	-0.6	-1.6	-1.4	-1.3
1.163	1.1		3.4	-0.1	3.2	2.9	1.9
-3.13							
1.71							

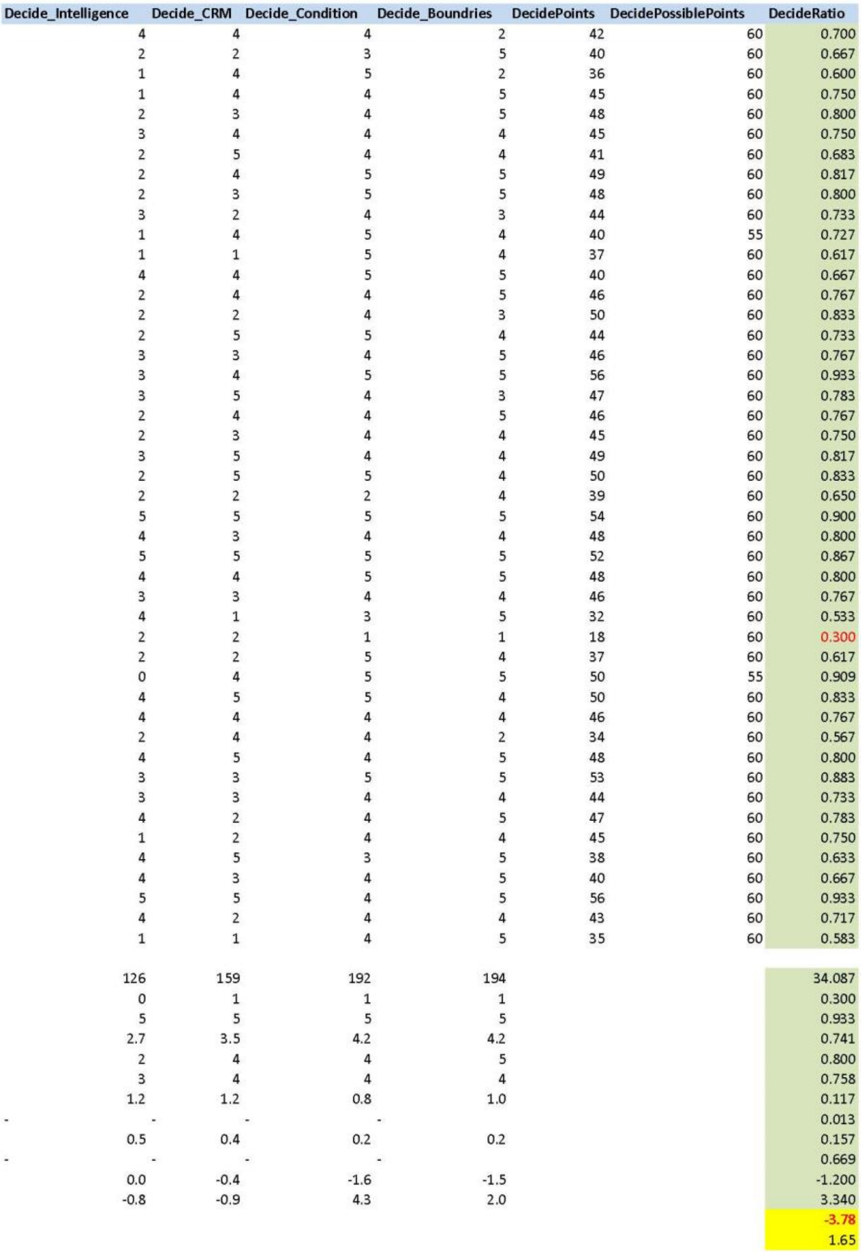


	Connect_Educate	Connect_Buying	Connect_Activities	Connect_Resources	Connect_Protect	ConnectPoints	ConnectPossiblePoints
	2	3	2	3	3	35	60
	4	4	4	3	3	46	60
	5	4	3	5	2	48	60
	5	4	1	4	2	47	60
	4	3	3	2	4	45	60
	5	5	2	2	4	44	60
	5	2	4	3	4	41	60
	5	5	3	5	4	50	60
	4	3	2	4	3	44	60
	3	3	4	4	2	44	60
	5	2	2	1	4	40	60
	4	4	3	4	3	45	60
	5	1	4	1	3	42	60
	5	5	4	2	4	51	60
	4	4	4	4	3	48	60
	2	3	3	4	4	41	60
	5	5	4	4	5	51	60
	5	5	5	4	5	56	60
	5	4	5	2	3	47	60
	4	4	3	4	3	50	60
	3	2	3	5	4	45	60
	3	3	4	2	4	45	60
	5	4	5	2	5	56	60
	2	4	4	4	4	38	60
	5	2	2	2	5	49	60
	5	3	4	4	4	52	60
	4	3	3	3	3	43	60
	5	5	4	4	4	54	60
	4	4	4	4	4	48	60
	2	0	1	4	4	31	55
	3	1	1	3	1	24	60
	4	2	2	2	4	43	60
	5	5	5	5	2	56	60
	5	1	4	4	1	45	60
	5	4	2	2	5	47	60
	4	2	2	4	4	33	60
	4	2	4	2	5	41	60
	4	5	4	2	5	53	60
	4	3	3	3	3	44	60
	5	4	3	2	4	48	60
	4	2	4	2	1	38	60
	1	5	2	2	2	38	60
	5	4	2	2	4	45	60
	5	3	3	3	4	47	60
	5	3	3	1	3	44	60
	4	2	2	2	3	32	60
	191	151	145	140	160		
	1	0	1	1	1		
	5	5	5	5	5		
	4.2	3.3	3.2	3.0	3.5		
	5	4	4	4	4		
	4	3	3	3	4		
	1.1	1.3	1.1	1.2	1.1		
-	-	-	-	-			
	0.3	0.4	0.3	0.4	0.3		
-	-	-	-	-			
	-1.3	-0.4	-0.2	0.0	-0.6		
	1.0	-0.5	-0.8	-1.2	0.0		



Learn_Offsite	Learn_Portal	Learn_BI	LearnPoints	LearnPossiblePoints	LearnRatio	LearnInteger	Learn Z-SCORE	Connect_Beliefs
3	5	5	32	45	0.711	71		3
3	4	5	39	45	0.867	87		4
1	4	1	29	45	0.644	64		4
3	5	5	34	45	0.756	76		5
4	4	4	35	45	0.778	78		3
4	5	2	39	45	0.867	87		4
2	4	2	29	45	0.644	64		4
5	4	5	41	45	0.911	91		4
4	5	2	34	45	0.756	76		4
5	1	4	29	45	0.644	64		4
4	4	1	31	45	0.689	69		4
4	4	2	26	45	0.578	58		4
3	4	2	30	45	0.667	67		4
4	5	4	40	45	0.889	89		5
2	4	4	31	45	0.689	69		4
4	5	3	31	45	0.689	69		4
5	5	5	37	45	0.822	82		4
4	5	5	42	45	0.933	93		4
2	1	4	30	45	0.667	67		4
5	5	2	41	45	0.911	91		5
4	2	2	30	45	0.667	67		5
4	4	4	34	45	0.756	76		4
5	5	2	40	45	0.889	89		5
3	2	2	22	45	0.489	49		4
2	5	2	33	45	0.733	73		5
4	4	5	34	40	0.850	85		5
3	4	4	40	45	0.889	89		4
2	5	2	33	45	0.733	73		5
4	4	3	35	45	0.778	78		4
3	4	5	28	45	0.622	62		3
2	1	1	10	45	0.222	22		4
3	3	2	35	45	0.778	78		4
4	4	5	37	45	0.822	82		5
4	5	4	35	45	0.778	78		5
3	5	5	39	45	0.867	87		4
2	4	4	24	45	0.533	53		2
2	5	5	29	45	0.644	64		3
5	5	4	44	45	0.978	98		5
3	4	3	30	45	0.667	67		4
5	2	2	32	45	0.711	71		4
2	4	2	32	45	0.711	71		5
3	5	5	35	45	0.778	78		5
3	4	4	32	45	0.711	71		4
3	2	4	36	45	0.800	80		4
4	4	5	35	45	0.778	78		4
3	4	2	32	45	0.711	71		4
156	183	155			34.006	3402		191
1	1	1			0.222	22		2
5	5	5			0.978	98		5
3.4	4.0	3.4			0.739	74.0		4.2
4	4	2			0.778	78		4
3	4	4			0.744	74.5		4
1.0	1.2	1.4			0.132	13.3		0.7
-	-	-			0.017	173.4		-
0.3	0.3	0.4			0.179	0.2		0.2
-	-	-			0.596	-		-
-0.1	-1.4	-0.2			-1.234	-1.2		-0.7
-0.7	1.2	-1.5			4.000	4.0		1.3
					-3.90			
					1.80			

DecideInteger	Decide Z-SCORE	Learn_Training	Learn_Mentor	Learn_Reimburse	Learn_Priority	Learn_Capture	Learn_Venue
70		3	3	5	4	2	2
67		5	4	5	5	4	4
60		4	4	5	3	4	3
75		4	4	4	4	2	3
80		5	4	5	3	2	4
75		5	5	5	5	4	4
68		4	3	5	4	3	2
82		5	5	5	5	3	4
80		4	5	5	3	3	3
73		5	4	4	2	2	2
73		5	3	4	4	2	4
62		4	2	3	3	2	2
67		3	4	4	4	2	4
77		4	5	5	4	4	5
83		4	4	4	3	2	4
73		3	3	5	3	3	2
77		5	5	1	2	4	5
93		5	5	5	4	4	5
78		5	5	4	4	1	4
77		5	5	5	5	5	4
75		4	4	5	4	2	3
82		4	4	4	3	3	4
83		4	5	5	5	5	4
65		2	2	3	3	2	3
90		5	5	5	5	2	2
80		4	5	5	4	3	0
87		5	5	5	5	4	5
80		4	4	4	4	4	4
77		4	4	4	4	4	4
53		3	2	4	4	2	1
30		1	1	1	1	1	1
62		5	5	5	4	4	4
91		5	5	1	5	4	4
83		5	5	2	3	2	5
77		5	5	3	4	5	4
57		2	2	4	2	2	2
80		5	2	2	4	2	2
88		5	5	5	5	5	5
73		3	4	3	4	3	3
78		5	5	4	4	2	3
75		4	4	4	4	4	4
63		4	5	5	4	2	2
67		4	4	2	4	4	3
93		5	5	5	5	4	3
72		4	4	4	2	4	4
59		4	4	5	2	4	4
3410		192	187	187	172	141	153
30		1	1	1	1	1	0
93		5	5	5	5	5	5
74.1		4.2	4.1	4.1	3.7	3.1	3.3
80		5	5	5	4	2	4
76		4	4	4	4	3	4
11.6		0.9	1.1	1.2	1.0	1.1	1.2
131.5		-	-	-	-	-	-
0.2		0.2	0.3	0.3	0.3	0.4	0.4
-		-	-	-	-	-	-
-1.2		-1.3	-1.1	-1.3	-0.7	0.1	-0.7
3.5		1.9	0.5	1.0	0.1	-1.2	0.0



Decide_Intelligence	Decide_CRM	Decide_Condition	Decide_Boundries	DecidePoints	DecidePossiblePoints	DecideRatio
4	4	4	2	42	60	0.700
2	2	3	5	40	60	0.667
1	4	5	2	36	60	0.600
1	4	4	5	45	60	0.750
2	3	4	5	48	60	0.800
3	4	4	4	45	60	0.750
2	5	4	4	41	60	0.683
2	4	5	5	49	60	0.817
2	3	5	5	48	60	0.800
3	2	4	3	44	60	0.733
1	4	5	4	40	55	0.727
1	1	5	4	37	60	0.617
4	4	5	5	40	60	0.667
2	4	4	5	46	60	0.767
2	2	4	3	50	60	0.833
2	5	5	4	44	60	0.733
3	3	4	5	46	60	0.767
3	4	5	5	56	60	0.933
3	5	4	3	47	60	0.783
2	4	4	5	46	60	0.767
2	3	4	4	45	60	0.750
3	5	4	4	49	60	0.817
2	5	5	4	50	60	0.833
2	2	2	4	39	60	0.650
5	5	5	5	54	60	0.900
4	3	4	4	48	60	0.800
5	5	5	5	52	60	0.867
4	4	5	5	48	60	0.800
3	3	4	4	46	60	0.767
4	1	3	5	32	60	0.533
2	2	1	1	18	60	0.300
2	2	5	4	37	60	0.617
0	4	5	5	50	55	0.909
4	5	5	4	50	60	0.833
4	4	4	4	46	60	0.767
2	4	4	2	34	60	0.567
4	5	4	5	48	60	0.800
3	3	5	5	53	60	0.883
3	3	4	4	44	60	0.733
4	2	4	5	47	60	0.783
1	2	4	4	45	60	0.750
4	5	3	5	38	60	0.633
4	3	4	5	40	60	0.667
5	5	4	5	56	60	0.933
4	2	4	4	43	60	0.717
1	1	4	5	35	60	0.583
126	159	192	194			34.087
0	1	1	1			0.300
5	5	5	5			0.933
2.7	3.5	4.2	4.2			0.741
2	4	4	5			0.800
3	4	4	4			0.758
1.2	1.2	0.8	1.0			0.117
-	-	-	-			0.013
0.5	0.4	0.2	0.2			0.157
-	-	-	-			0.669
0.0	-0.4	-1.6	-1.5			-1.200
-0.8	-0.9	4.3	2.0			3.340
						-3.78
						1.65

APPENDIX III: Understanding of Data

This appendix holds statistical information about the data used in the research.

Graphical representation of answers combined into competence areas, organizational resilience and combined seven competence areas:

Note: In Fig. A3.1 – A.3.9 horizontal X-axis represents the value of 'ORInteger' column, which is the sum of points received in the 'OR Area'. The vertical axis represents the ratio measured as the number of 'points' within competence area divided by the total number of possible points for that competence area.

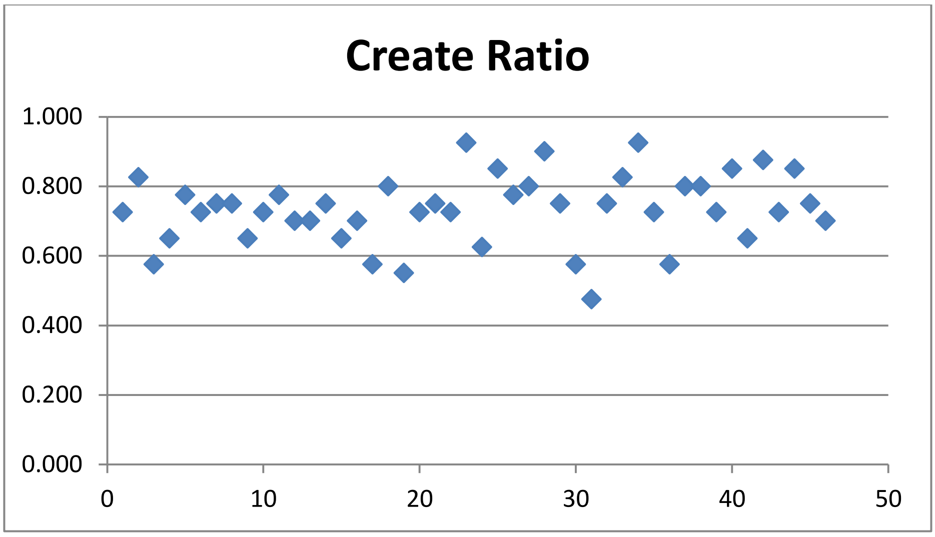


Fig. A3.1: Create ratio vs. OR Integer

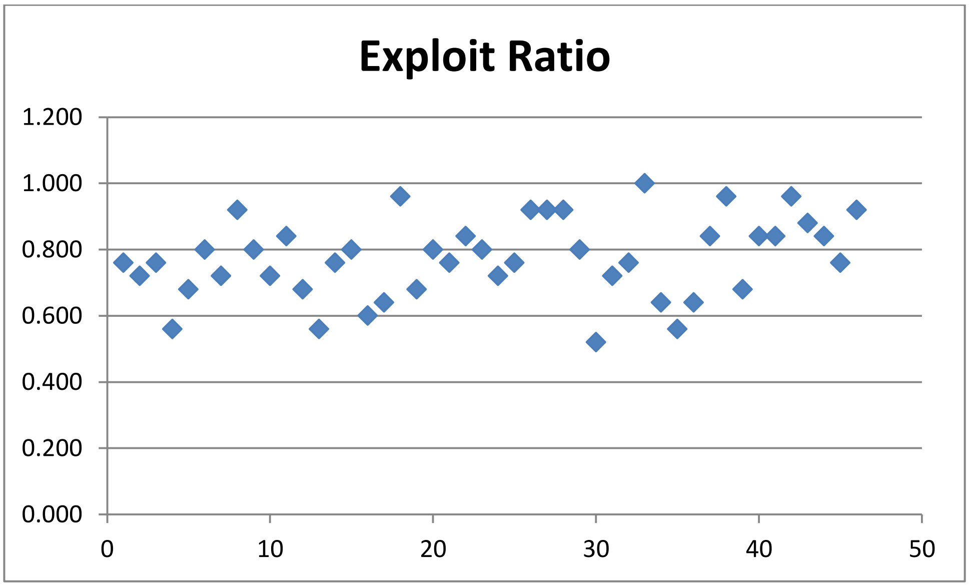


Fig. A3.2: Exploit ratio vs. OR Integer

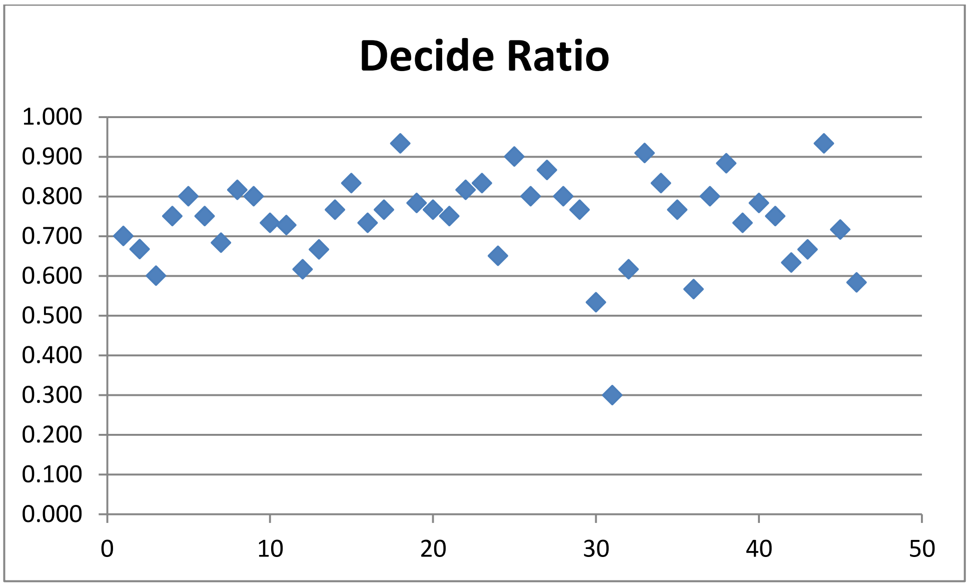


Fig. A3.3: Decide ratio vs. OR Integer

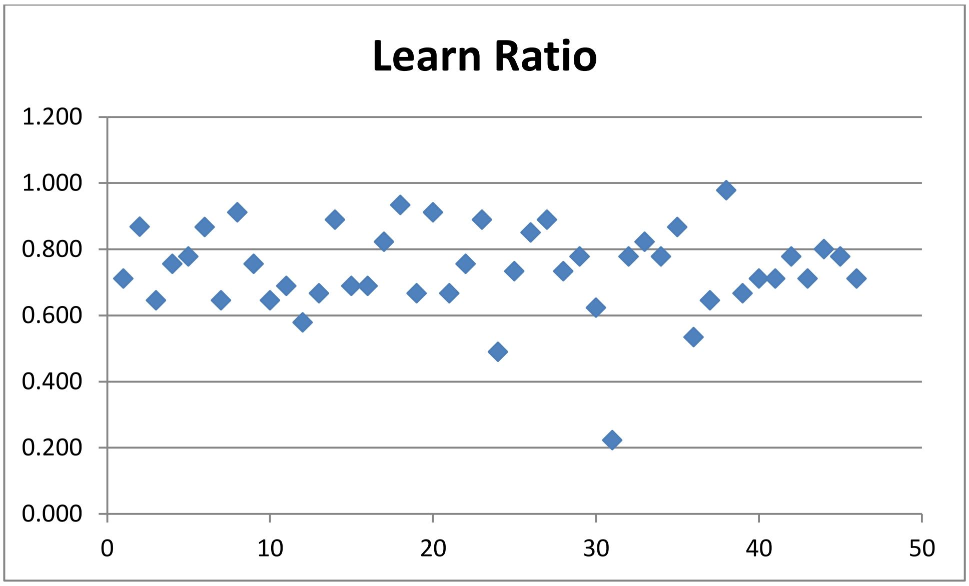


Fig. A3.4: Learn ratio vs. OR Integer

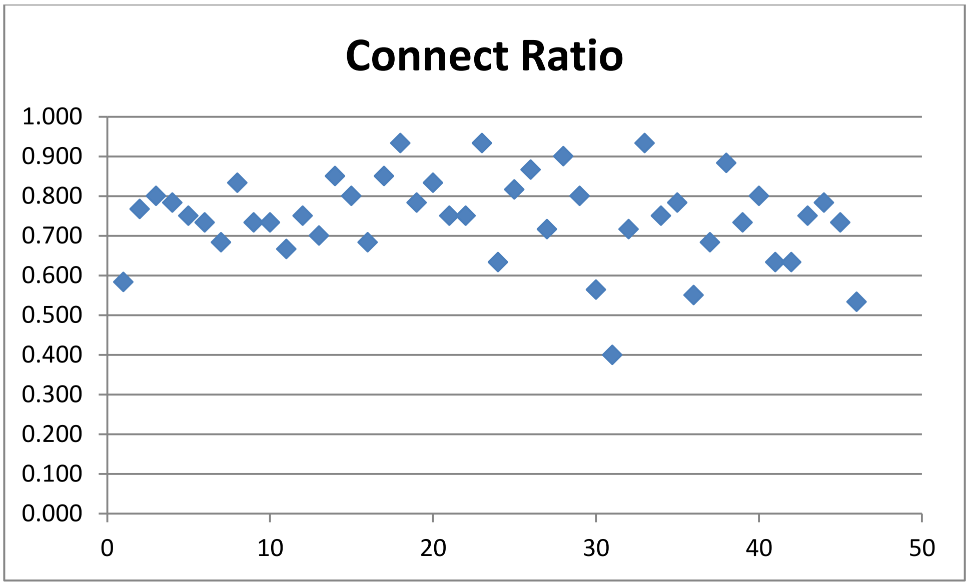


Fig. A3.5: Connect ratio vs. OR Integer

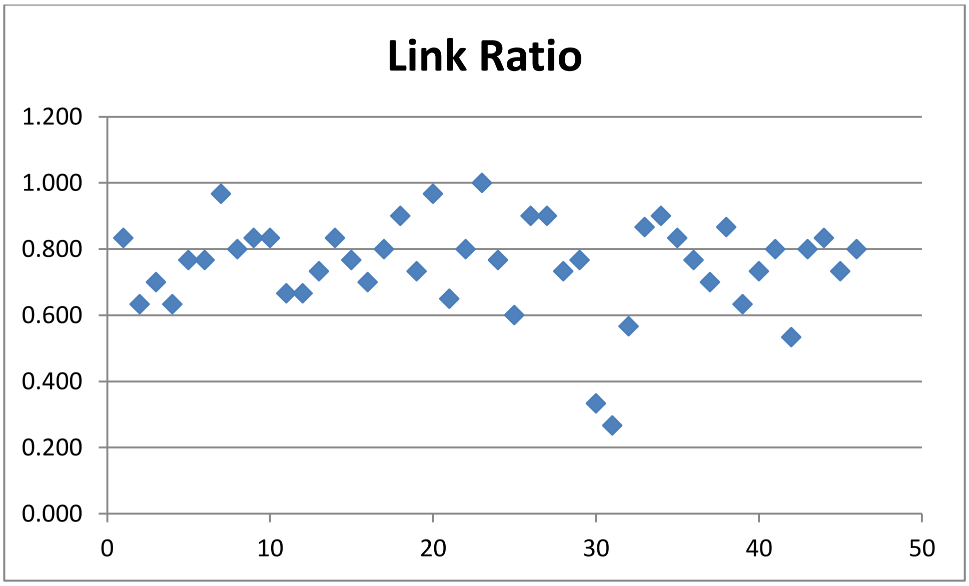


Fig. A3.6: Link ratio vs. OR Integer

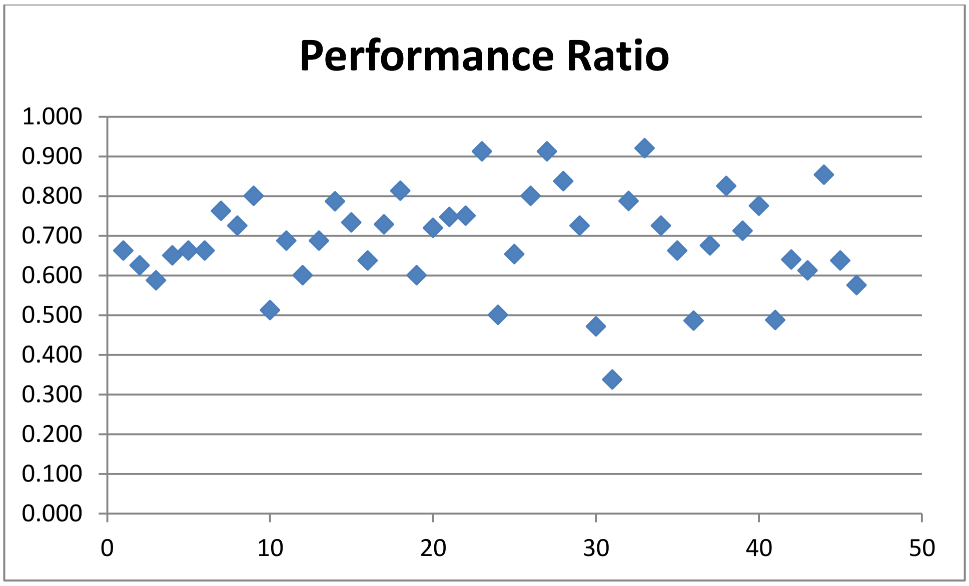


Fig. A3.7: Performance ratio vs. OR Integer

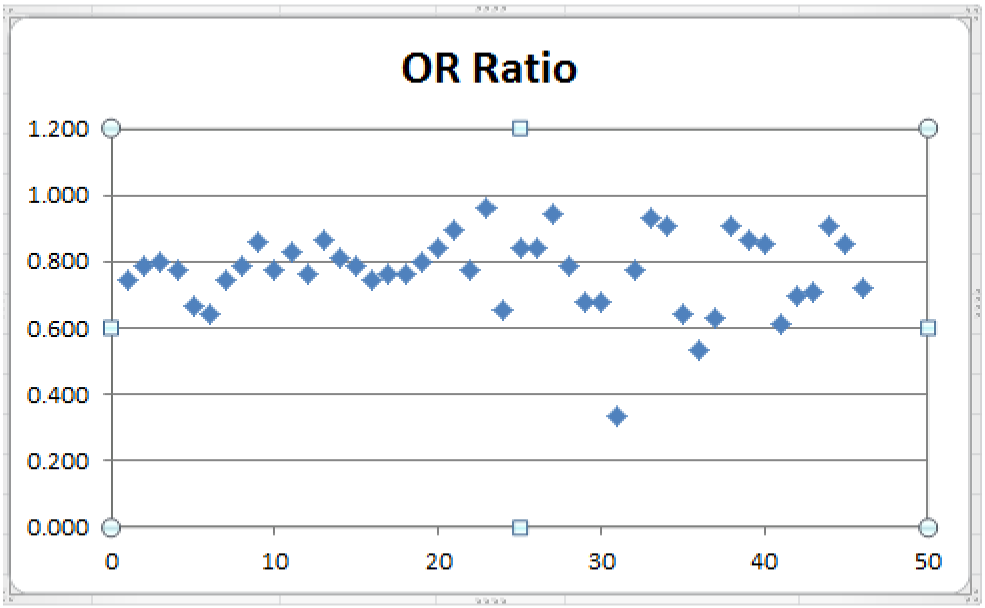


Fig. A3.8: OR ratio vs. OR Integer

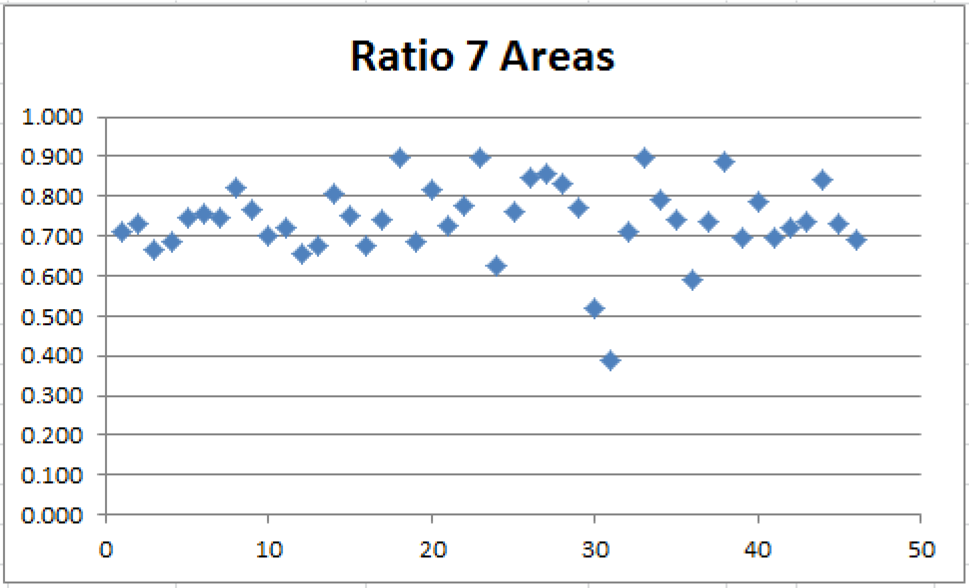
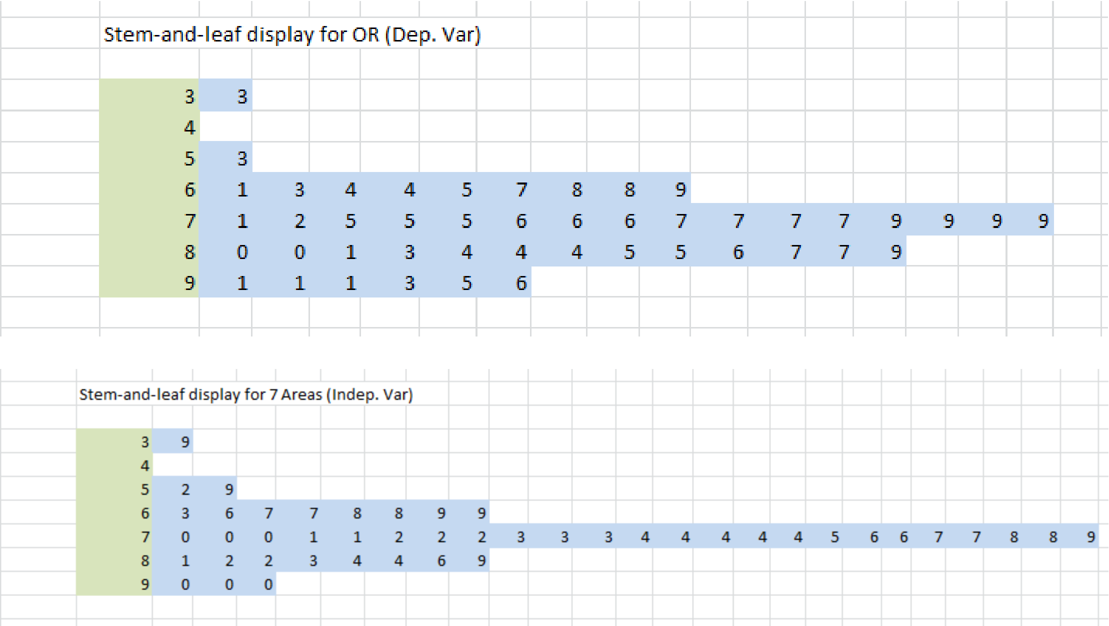


Fig. A3.9: seven Areas ratio vs. OR Integer

Stem and leaf representation for organizational resilience and combined seven competence areas:



Stem-and-leaf display for OR (Dep. Var)

3	3															
4																
5	3															
6	1	3	4	4	5	7	8	8	9							
7	1	2	5	5	5	6	6	6	7	7	7	7	9	9	9	9
8	0	0	1	3	4	4	4	5	5	6	7	7	9			
9	1	1	1	3	5	6										

Stem-and-leaf display for 7 Areas (Indep. Var)

3	9																							
4																								
5	2	9																						
6	3	6	7	7	8	8	9	9																
7	0	0	0	1	1	2	2	2	3	3	3	4	4	4	4	4	5	6	6	7	7	8	8	9
8	1	2	2	3	4	4	6	9																
9	0	0	0																					

Fig. A3.10: Stem and leaf display for OR

Additional analysis:

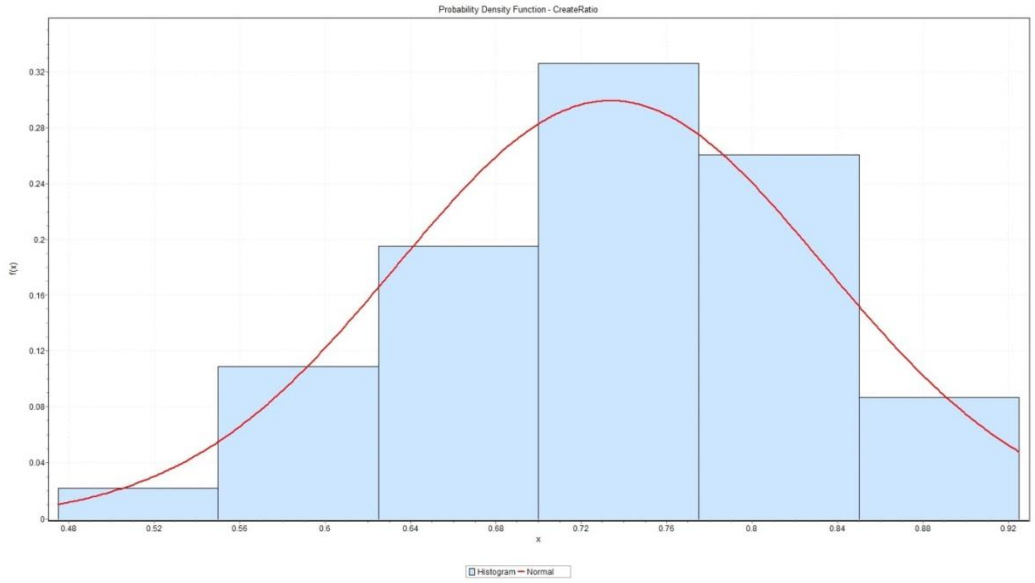


Fig. A3.11: Create ratio

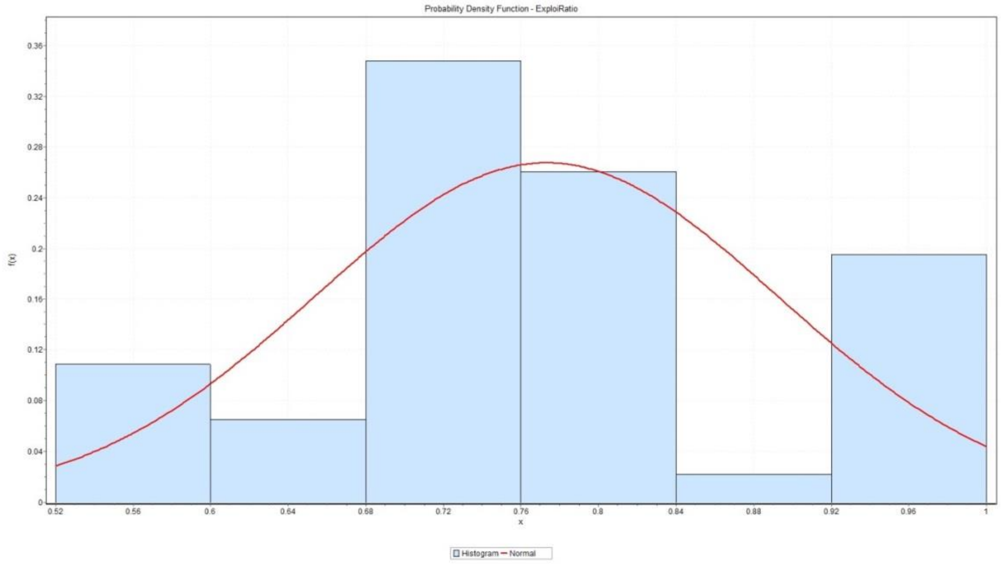


Fig. A3.12: Exploit ratio

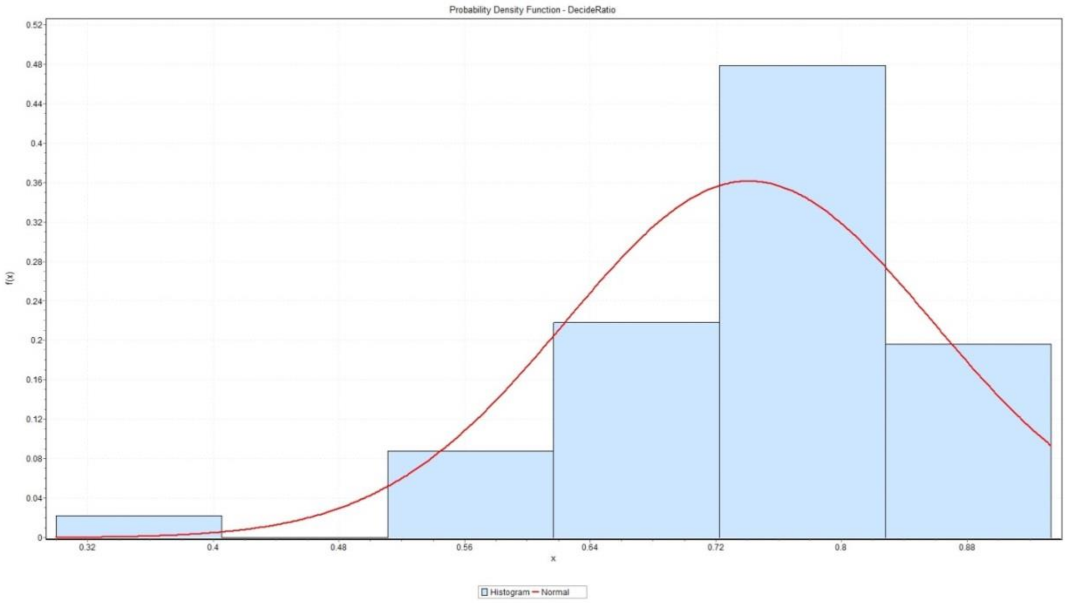


Fig. A3.13: Decide ratio

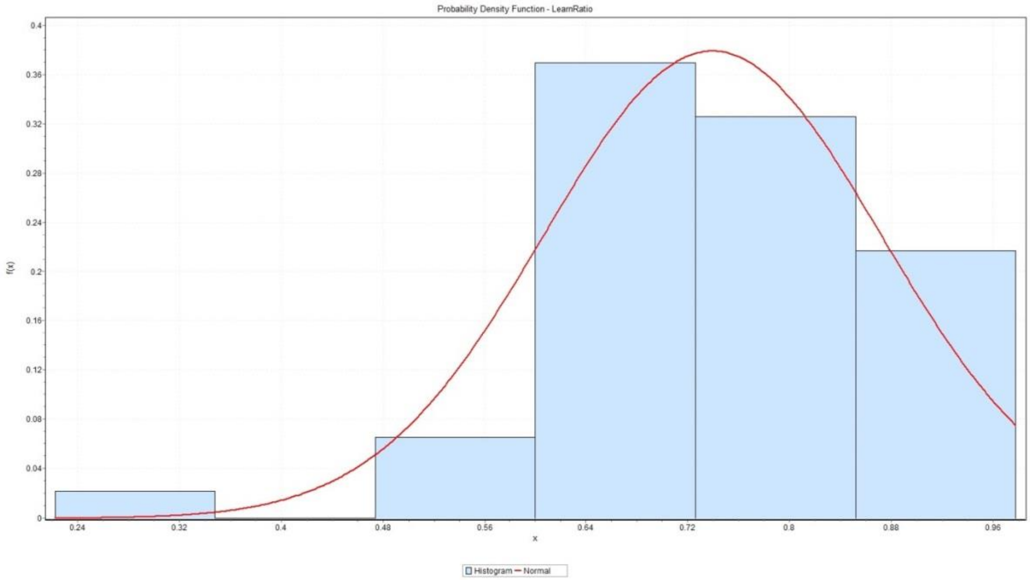


Fig. A3.14: Learn ratio

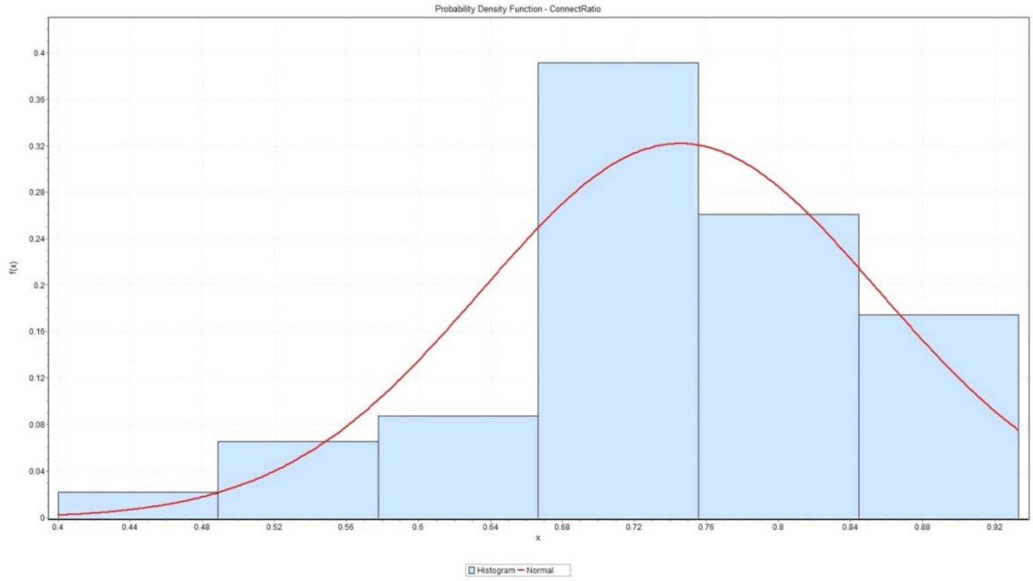


Fig. A3.15: Connect ratio

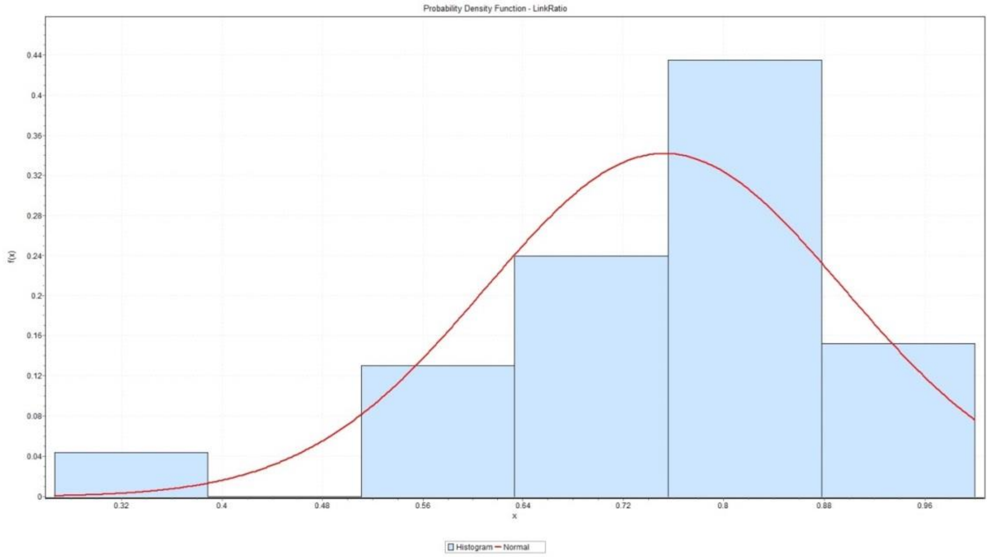


Fig. A3.16: Link ratio

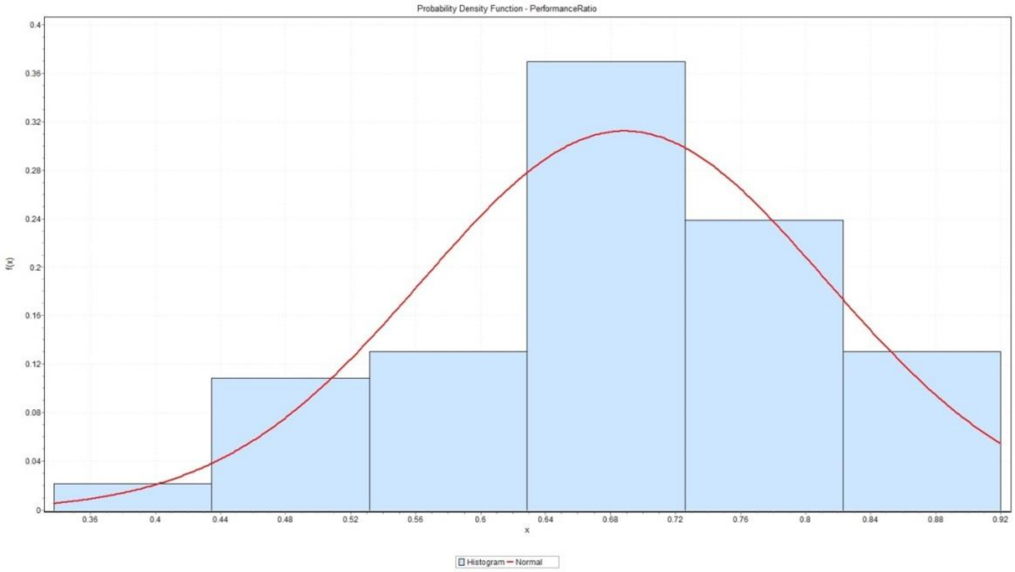


Fig. A3.17: Performance ratio

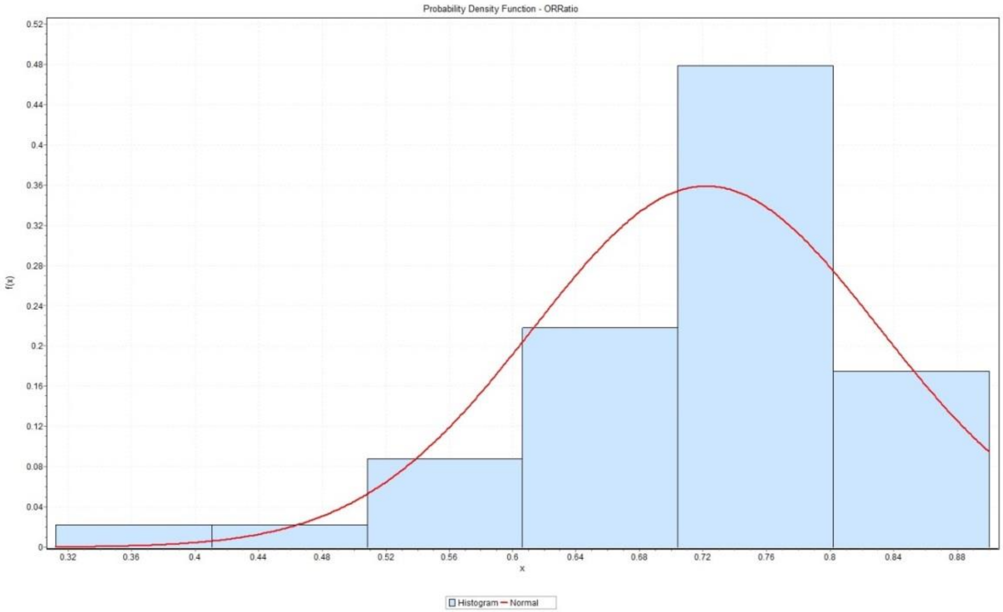


Fig. A3.18: OR ratio

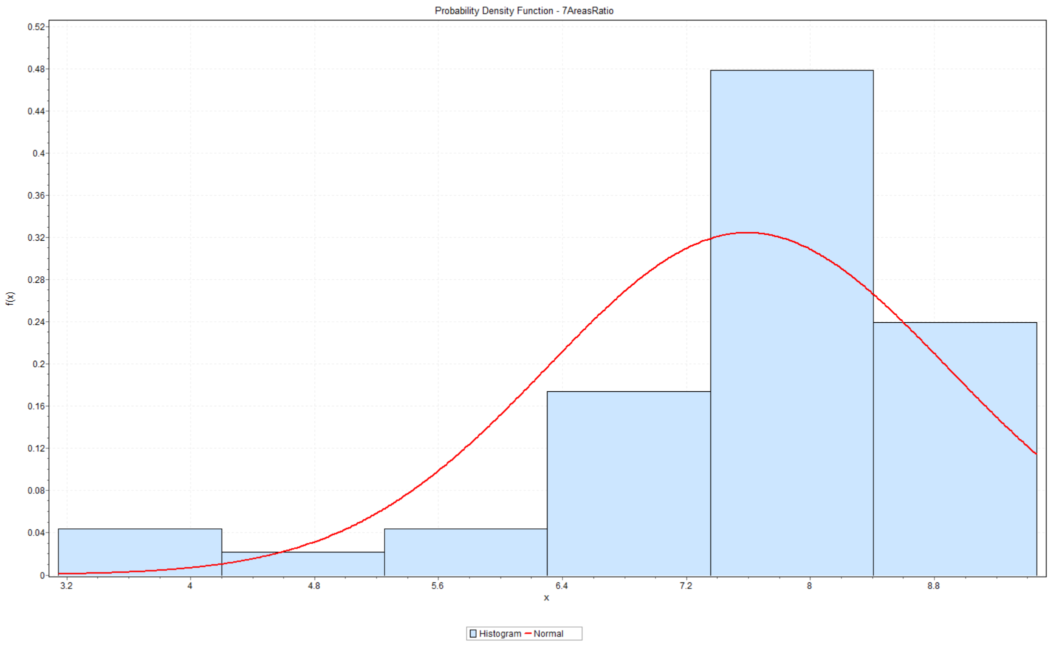
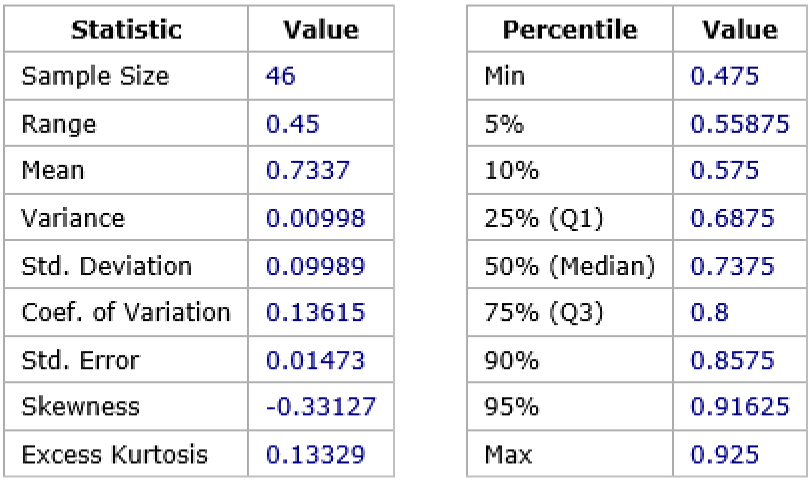


Fig. A3.19: Seven Areas ratio

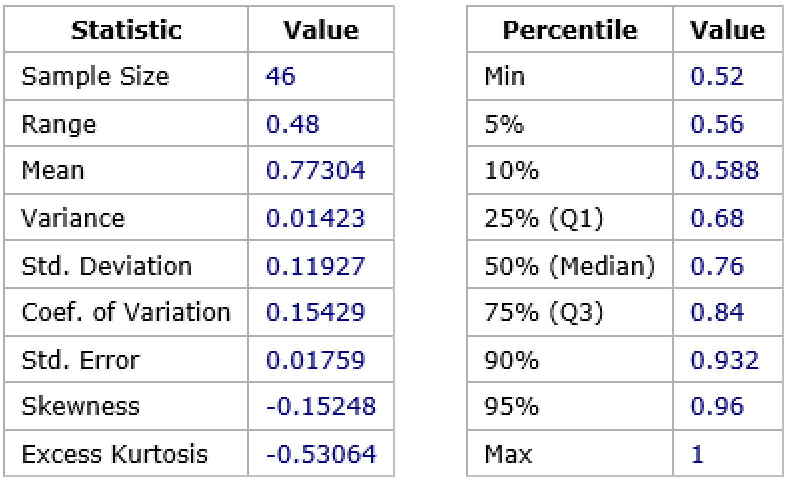
Descriptive Statistics.



Statistic	Value
Sample Size	46
Range	0.45
Mean	0.7337
Variance	0.00998
Std. Deviation	0.09989
Coef. of Variation	0.13615
Std. Error	0.01473
Skewness	-0.33127
Excess Kurtosis	0.13329

Percentile	Value
Min	0.475
5%	0.55875
10%	0.575
25% (Q1)	0.6875
50% (Median)	0.7375
75% (Q3)	0.8
90%	0.8575
95%	0.91625
Max	0.925

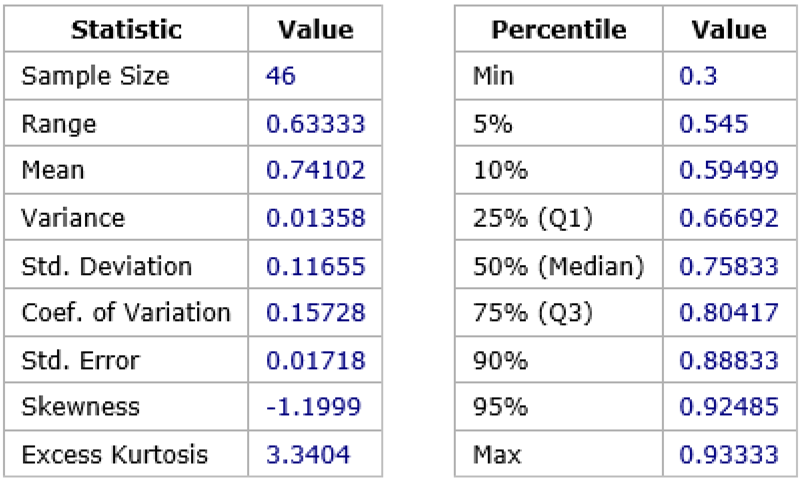
Fig. A3.20: Create competence statistics



Statistic	Value
Sample Size	46
Range	0.48
Mean	0.77304
Variance	0.01423
Std. Deviation	0.11927
Coef. of Variation	0.15429
Std. Error	0.01759
Skewness	-0.15248
Excess Kurtosis	-0.53064

Percentile	Value
Min	0.52
5%	0.56
10%	0.588
25% (Q1)	0.68
50% (Median)	0.76
75% (Q3)	0.84
90%	0.932
95%	0.96
Max	1

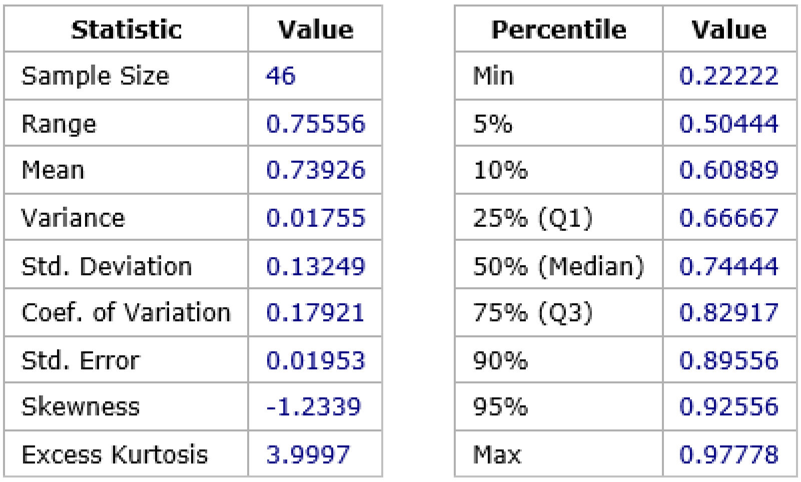
Fig. A3.21: Exploitation competence statistics



Statistic	Value
Sample Size	46
Range	0.63333
Mean	0.74102
Variance	0.01358
Std. Deviation	0.11655
Coef. of Variation	0.15728
Std. Error	0.01718
Skewness	-1.1999
Excess Kurtosis	3.3404

Percentile	Value
Min	0.3
5%	0.545
10%	0.59499
25% (Q1)	0.66692
50% (Median)	0.75833
75% (Q3)	0.80417
90%	0.88833
95%	0.92485
Max	0.93333

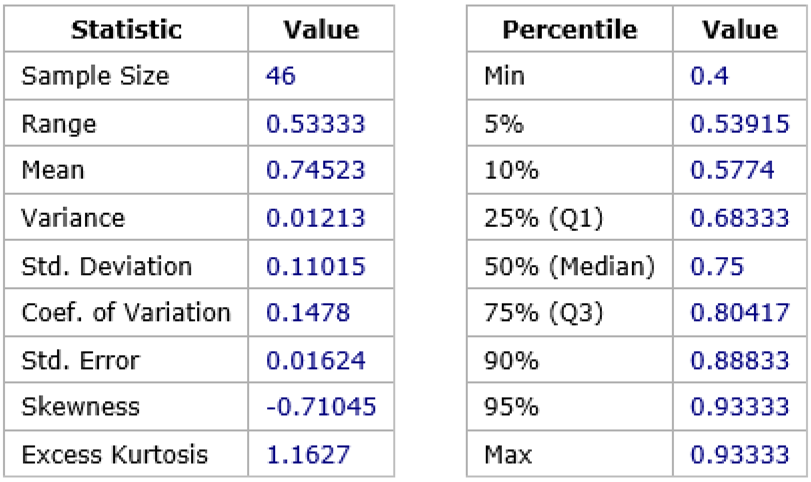
Fig. A3.22: Decide competence statistics



Statistic	Value
Sample Size	46
Range	0.75556
Mean	0.73926
Variance	0.01755
Std. Deviation	0.13249
Coef. of Variation	0.17921
Std. Error	0.01953
Skewness	-1.2339
Excess Kurtosis	3.9997

Percentile	Value
Min	0.22222
5%	0.50444
10%	0.60889
25% (Q1)	0.66667
50% (Median)	0.74444
75% (Q3)	0.82917
90%	0.89556
95%	0.92556
Max	0.97778

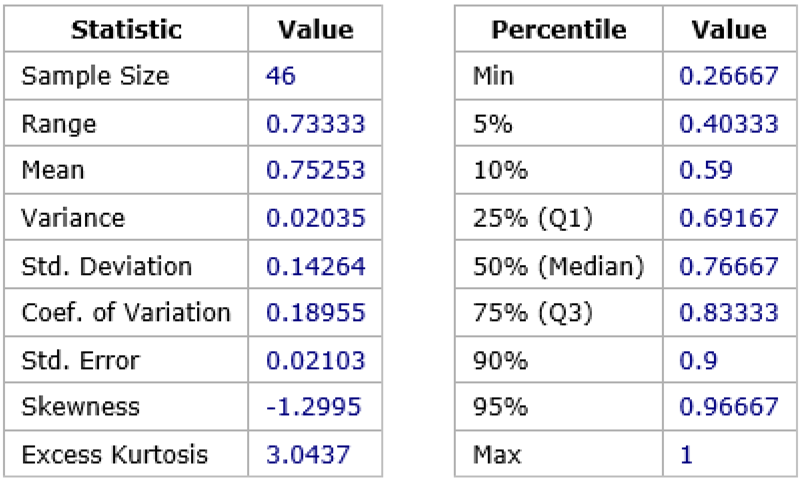
Fig. A3.23: Learn competence statistics



Statistic	Value
Sample Size	46
Range	0.53333
Mean	0.74523
Variance	0.01213
Std. Deviation	0.11015
Coef. of Variation	0.1478
Std. Error	0.01624
Skewness	-0.71045
Excess Kurtosis	1.1627

Percentile	Value
Min	0.4
5%	0.53915
10%	0.5774
25% (Q1)	0.68333
50% (Median)	0.75
75% (Q3)	0.80417
90%	0.88833
95%	0.93333
Max	0.93333

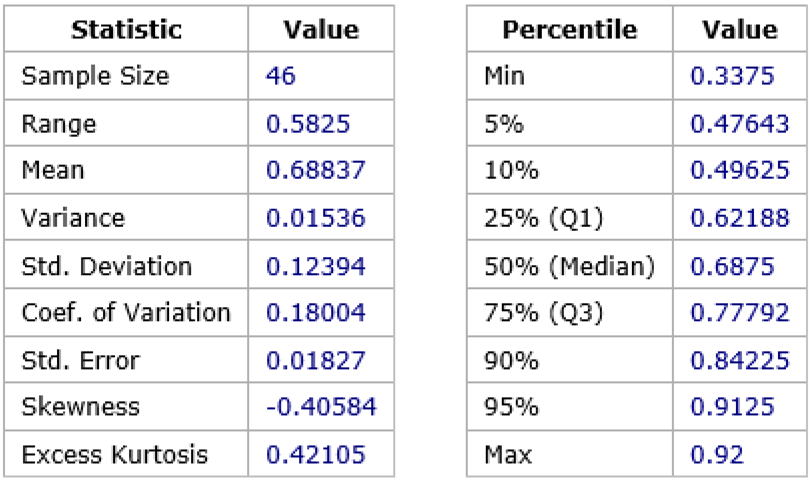
Fig. A3.24: Connect competence statistics



Statistic	Value
Sample Size	46
Range	0.73333
Mean	0.75253
Variance	0.02035
Std. Deviation	0.14264
Coef. of Variation	0.18955
Std. Error	0.02103
Skewness	-1.2995
Excess Kurtosis	3.0437

Percentile	Value
Min	0.26667
5%	0.40333
10%	0.59
25% (Q1)	0.69167
50% (Median)	0.76667
75% (Q3)	0.83333
90%	0.9
95%	0.96667
Max	1

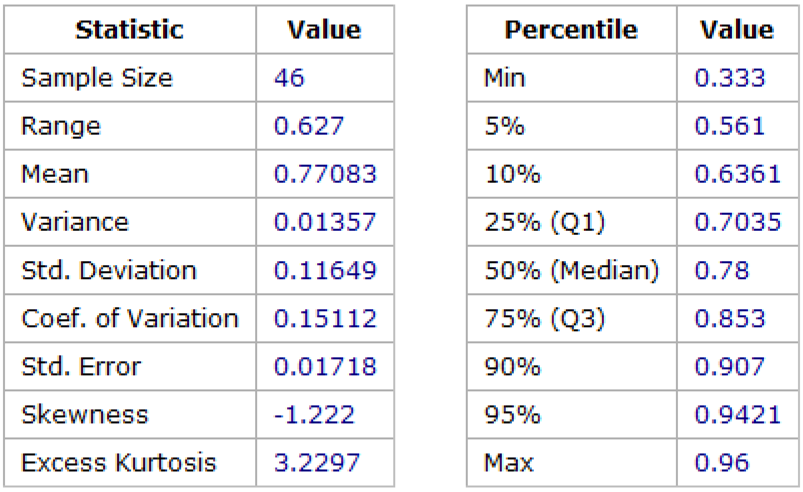
Fig. A3.25: Link competence statistics



Statistic	Value
Sample Size	46
Range	0.5825
Mean	0.68837
Variance	0.01536
Std. Deviation	0.12394
Coef. of Variation	0.18004
Std. Error	0.01827
Skewness	-0.40584
Excess Kurtosis	0.42105

Percentile	Value
Min	0.3375
5%	0.47643
10%	0.49625
25% (Q1)	0.62188
50% (Median)	0.6875
75% (Q3)	0.77792
90%	0.84225
95%	0.9125
Max	0.92

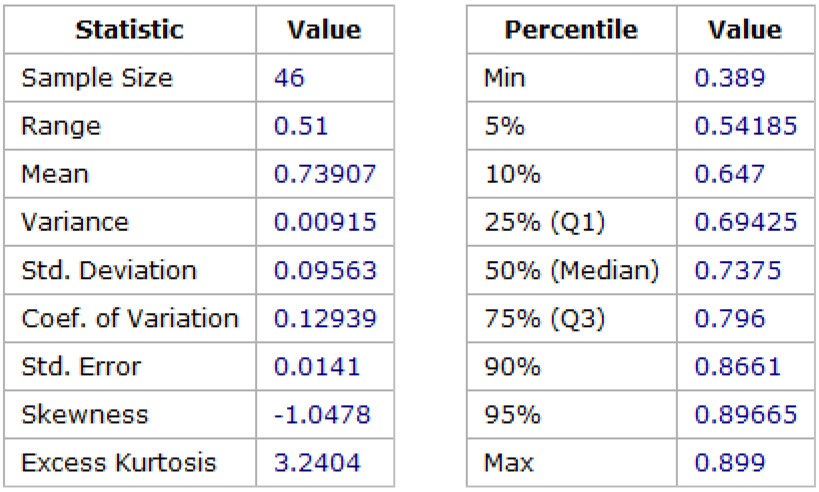
Fig. A3.26: Performance statistics



Statistic	Value
Sample Size	46
Range	0.627
Mean	0.77083
Variance	0.01357
Std. Deviation	0.11649
Coef. of Variation	0.15112
Std. Error	0.01718
Skewness	-1.222
Excess Kurtosis	3.2297

Percentile	Value
Min	0.333
5%	0.561
10%	0.6361
25% (Q1)	0.7035
50% (Median)	0.78
75% (Q3)	0.853
90%	0.907
95%	0.9421
Max	0.96

Fig. A3.27: Organizational resilience statistics



Statistic	Value
Sample Size	46
Range	0.51
Mean	0.73907
Variance	0.00915
Std. Deviation	0.09563
Coef. of Variation	0.12939
Std. Error	0.0141
Skewness	-1.0478
Excess Kurtosis	3.2404

Percentile	Value
Min	0.389
5%	0.54185
10%	0.647
25% (Q1)	0.69425
50% (Median)	0.7375
75% (Q3)	0.796
90%	0.8661
95%	0.89665
Max	0.899

Fig. A3.28: Seven competence areas statistics

Plots (Testing Association between Two Variables):

Note: The series of plots in Fig. A3.29 – A3.36 attempt to visually establish correlation between competence area and OR.

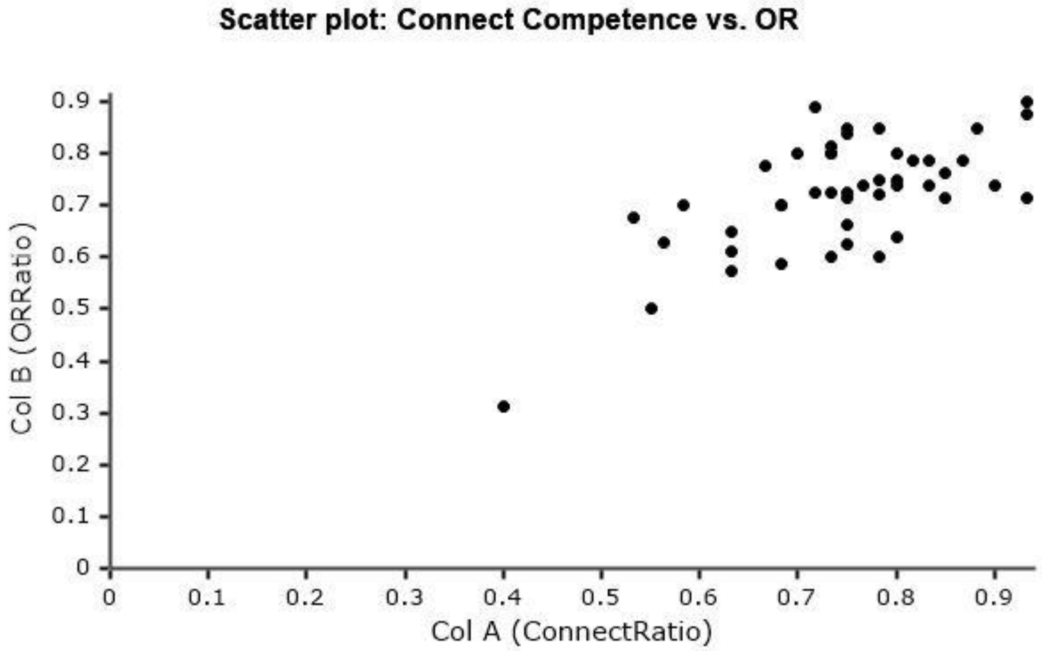


Fig. A3.29: Connect ratio vs. OR ratio

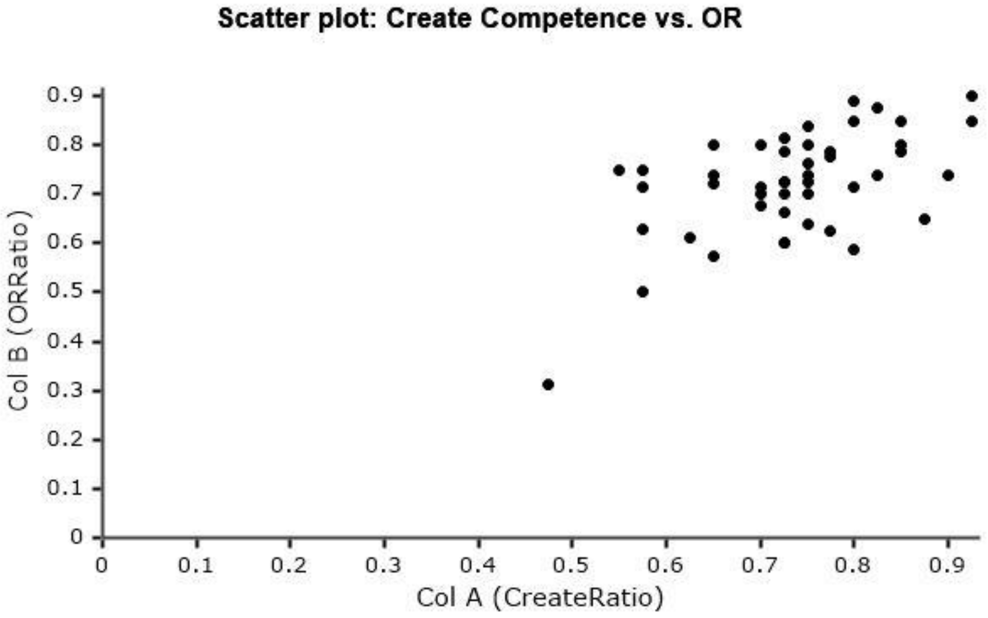


Fig. A3.30: Create ratio vs. OR ratio

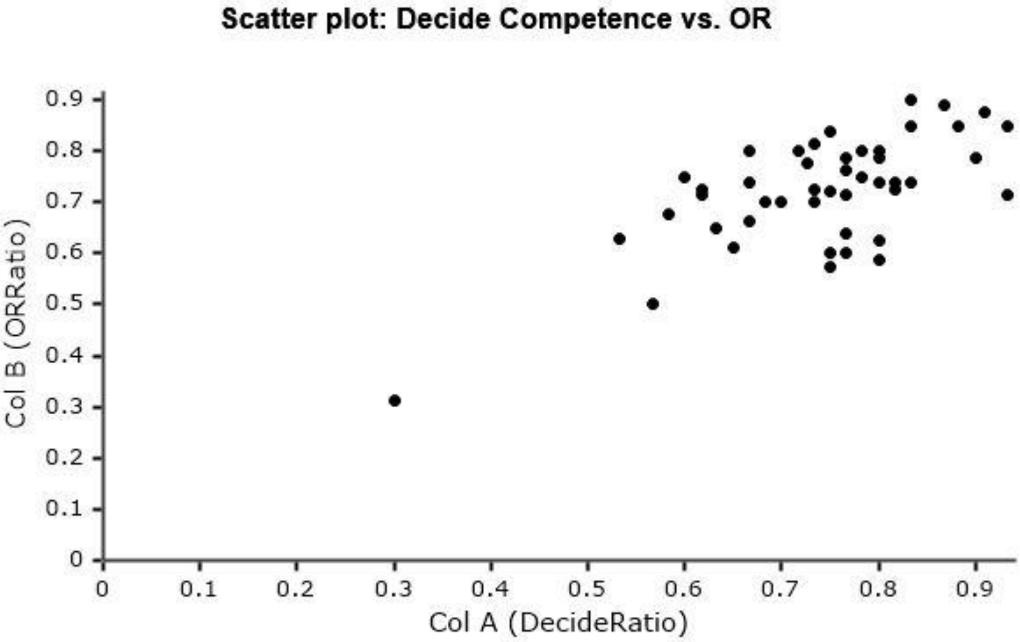


Fig. A3.31: Decide ratio vs. OR ratio

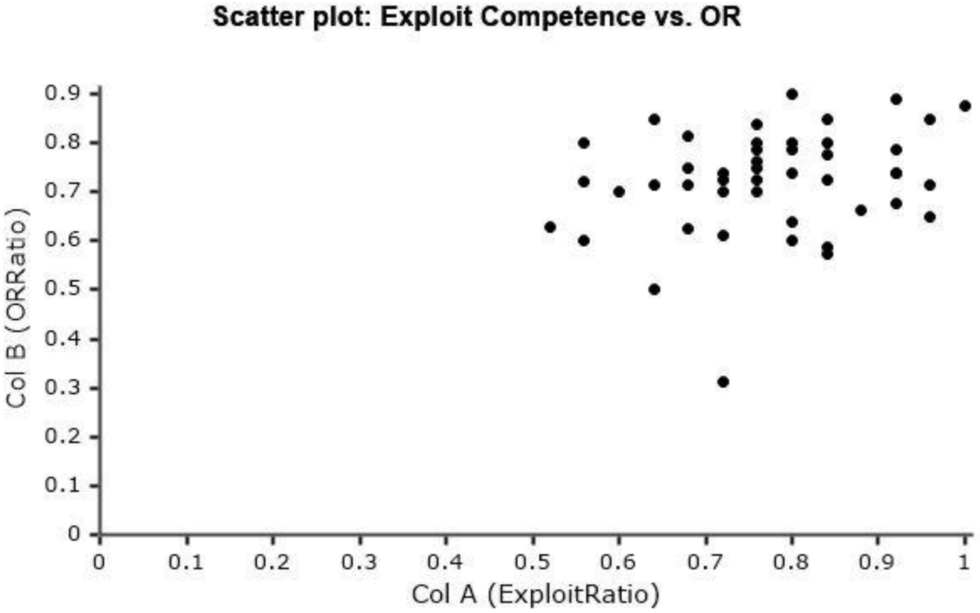


Fig. A3.32: Exploit ratio vs. OR ratio

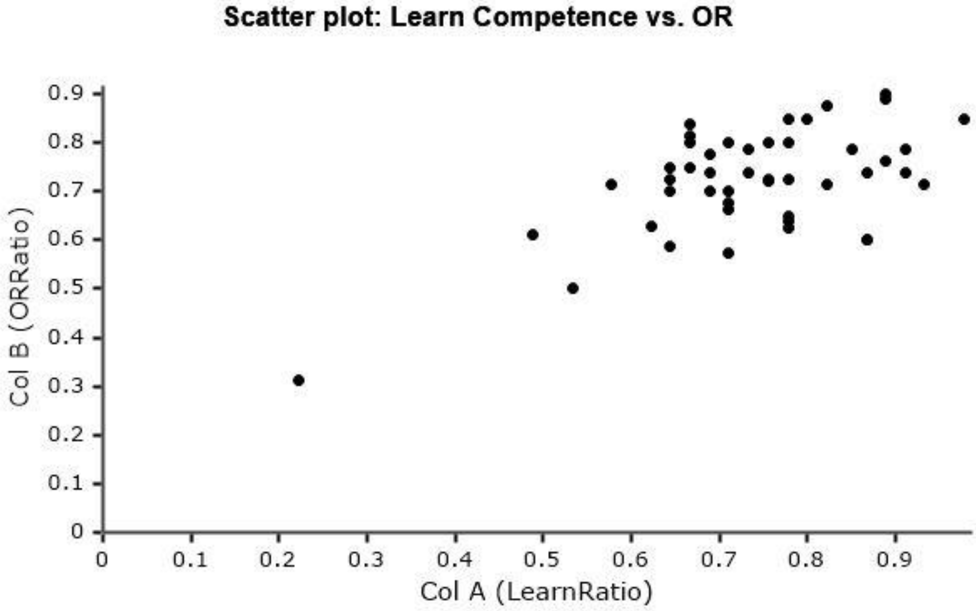


Fig. A3.33: Learn ratio vs. OR ratio

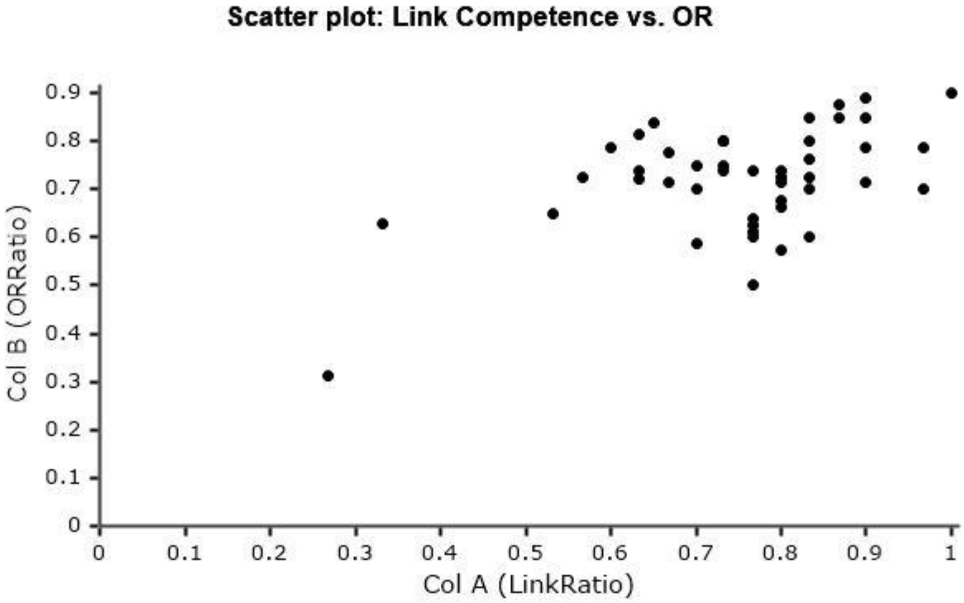


Fig. A3.34: Link ratio vs. OR ratio

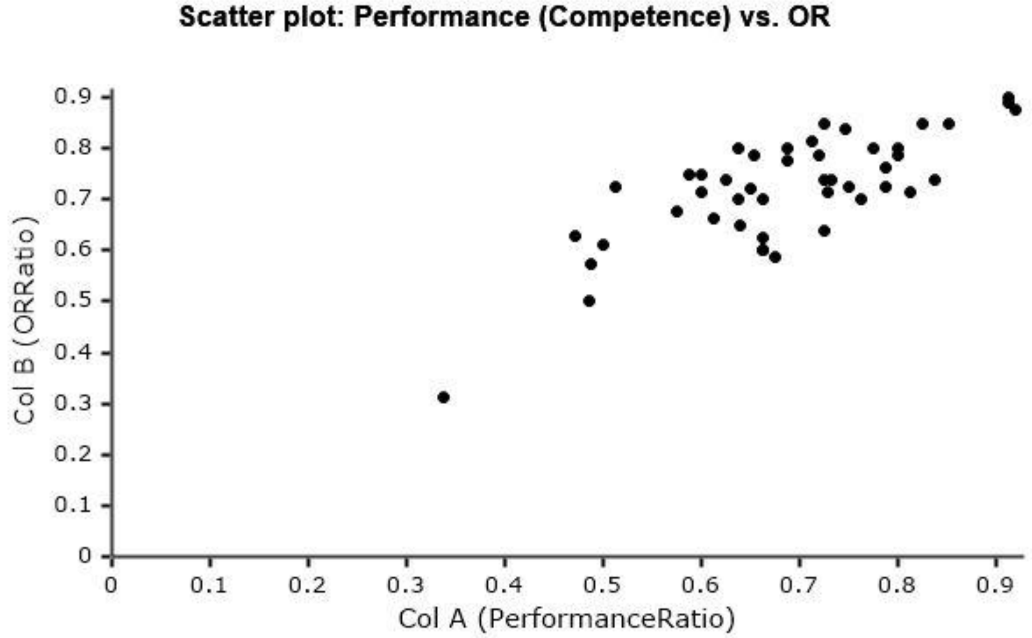


Fig. A3.35: Performance ratio vs. OR ratio

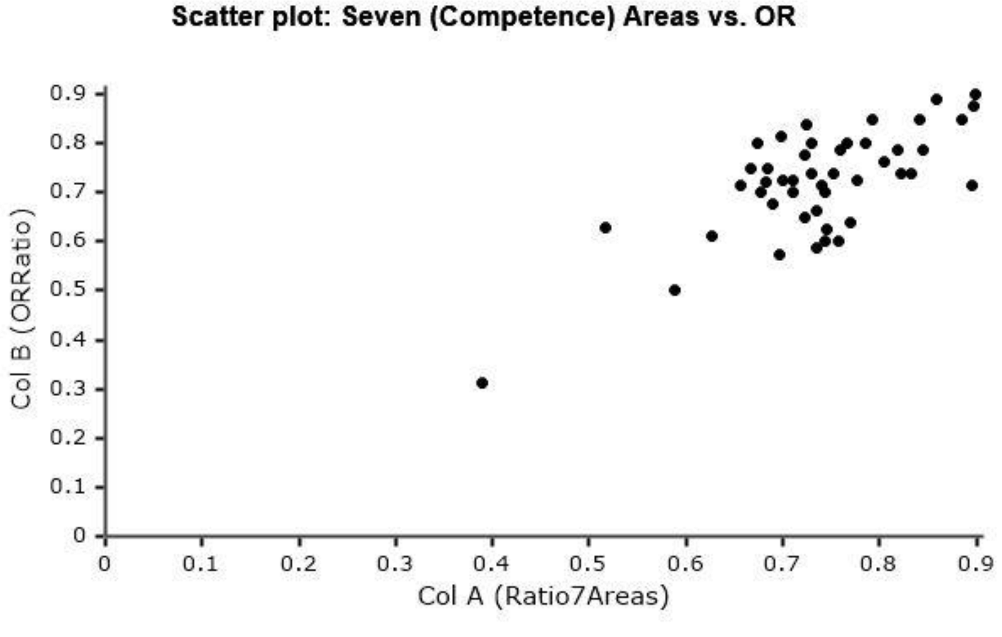


Fig. A3.36: Ratio 7 Areas vs. OR ratio

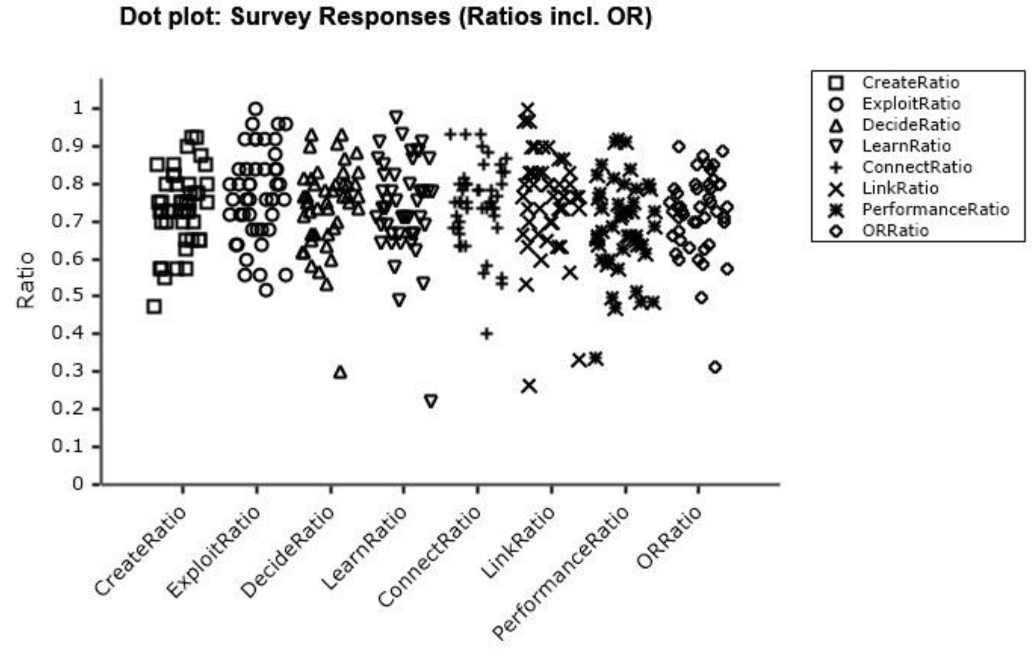


Fig. A3.37: Distribution of responses

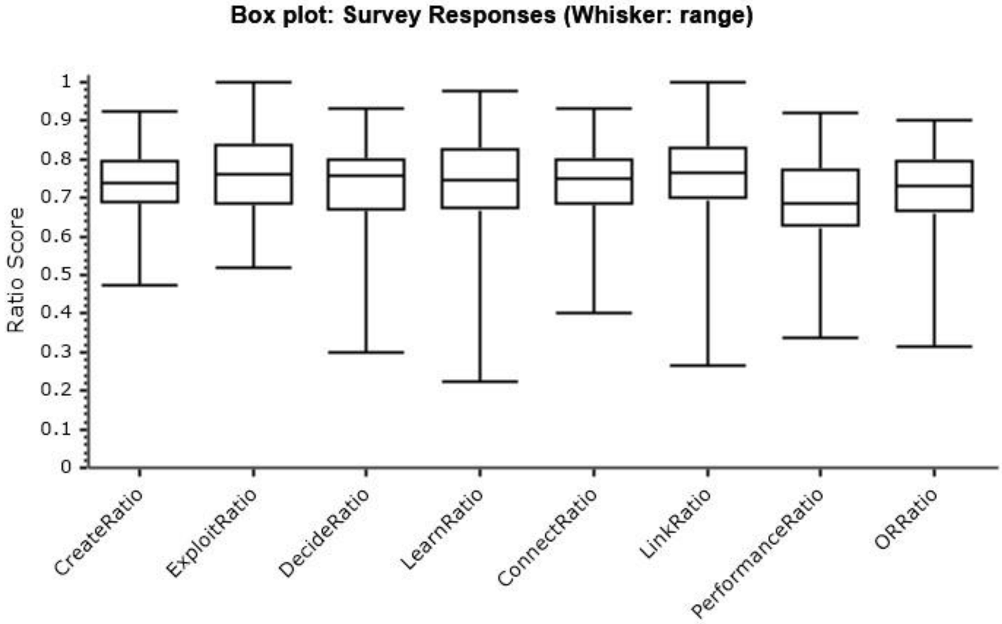
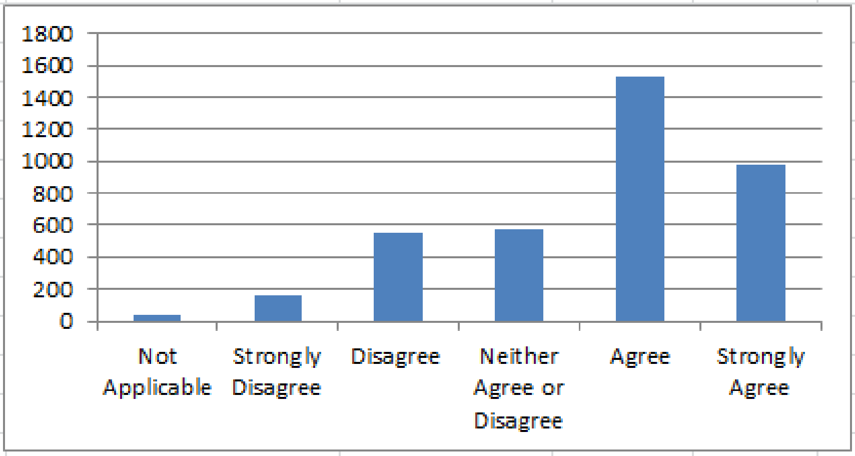


Fig. A3.38: Distribution of responses

Reply:	Points Assigned:	Count:	Percentage:
Not Applicable	0	34	0.9 %
Strongly Disagree	1	157	4.1 %
Disagree	2	549	14.4 %
Neither Agree or Disagree	3	570	14.9 %
Agree	4	1528	40.0 %
Strongly Agree	5	980	25.7 %
	Total:	3 818	100 %

Fig. A3.39: Distribution of replies



There are (84 - 1) questions answered by 46 firms. Total of 3 818 answered questions.

Fig. A3.40: Graphical representation of the distribution of replies

	Col A (Count:)
Data size (n)	6
Mean	636.333
Error	224.927
Standard deviation	550.957
C.I. (95%) of mean	± 578.194
Lower range	58.139

Upper range	1214.527
Minimum	34
Maximum	1528
Percentiles	
25th	126.25
50th	559.5
75th	1117
Coefficient of variation [%]	86.583
Geometric mean	368.436
Skewness	0.753
Kurtosis	0.06
Anderson-Darling test	
p-Value	0.9227
Pass normality test (p>0.05)?	Yes

Fig. A3.41: Analysis of un-grouped into sections/competence areas individual replies

Individual (Discrete) Questionnaire Answer Analysis.

The following section lists statistics related to individual answer (represented as a column in the Excel file containing input data).

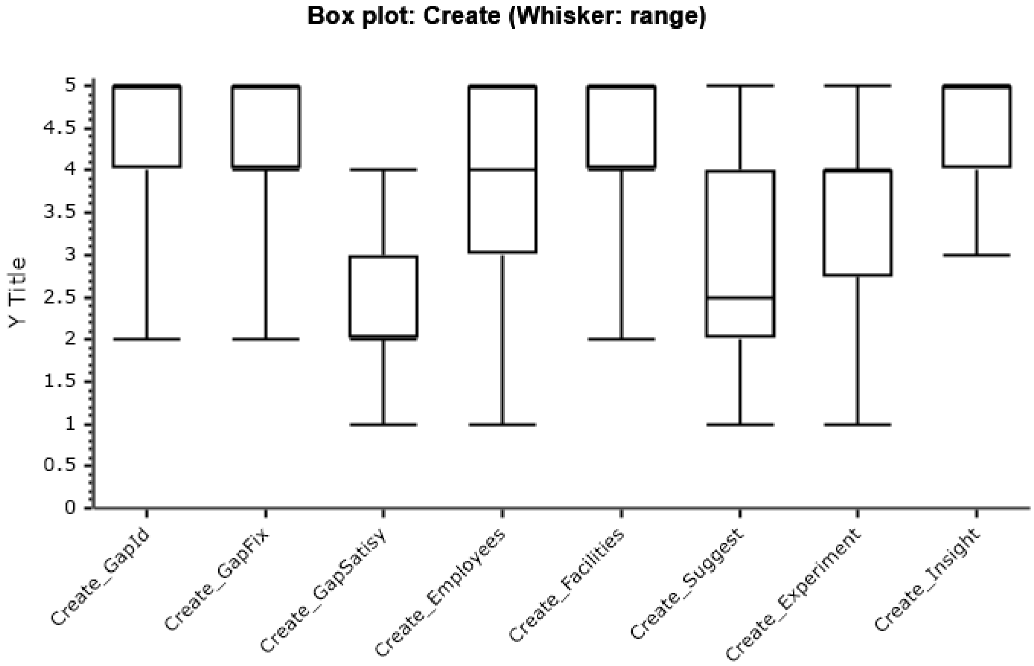


Fig. A3.42: Box plot of individual answers in the Create category

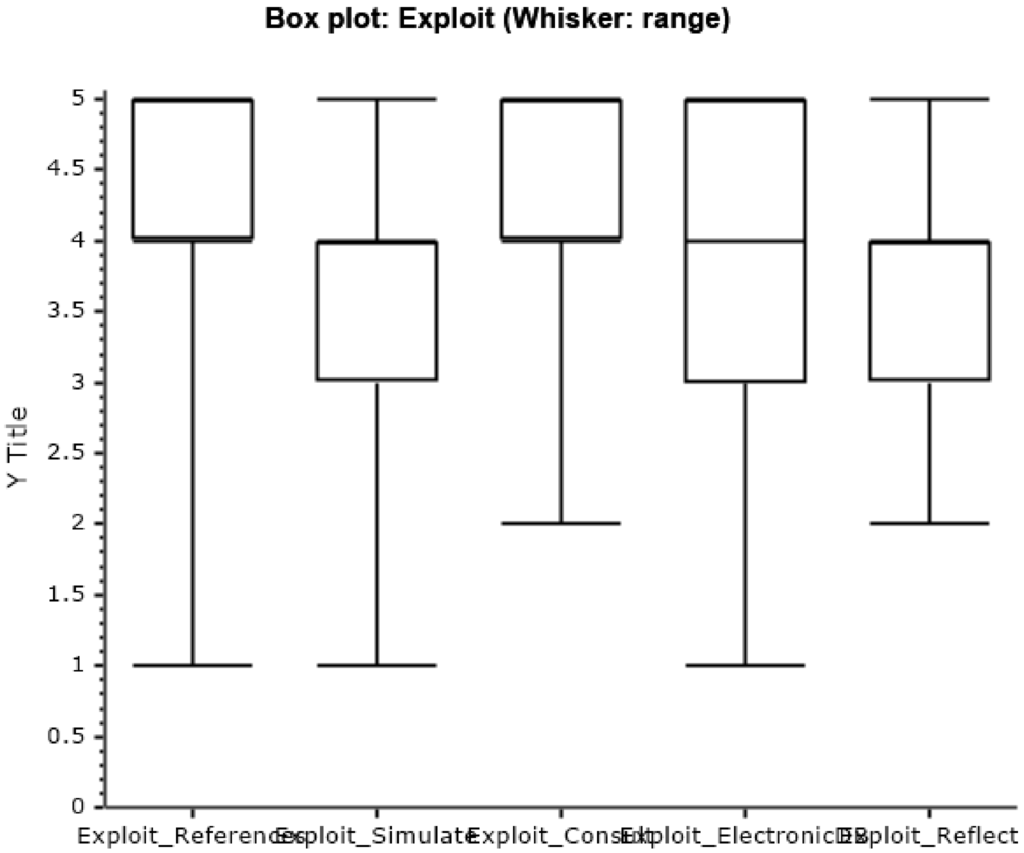


Fig. A3.43: Box plot of individual answers in the Exploit category

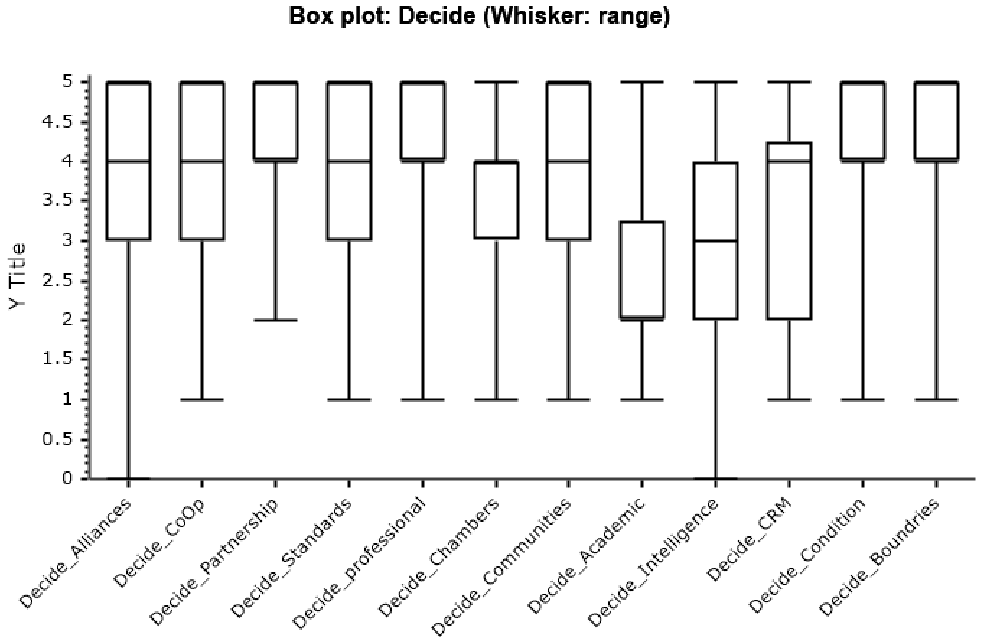


Fig. A3.44: Box plot of individual answers in the Decide category

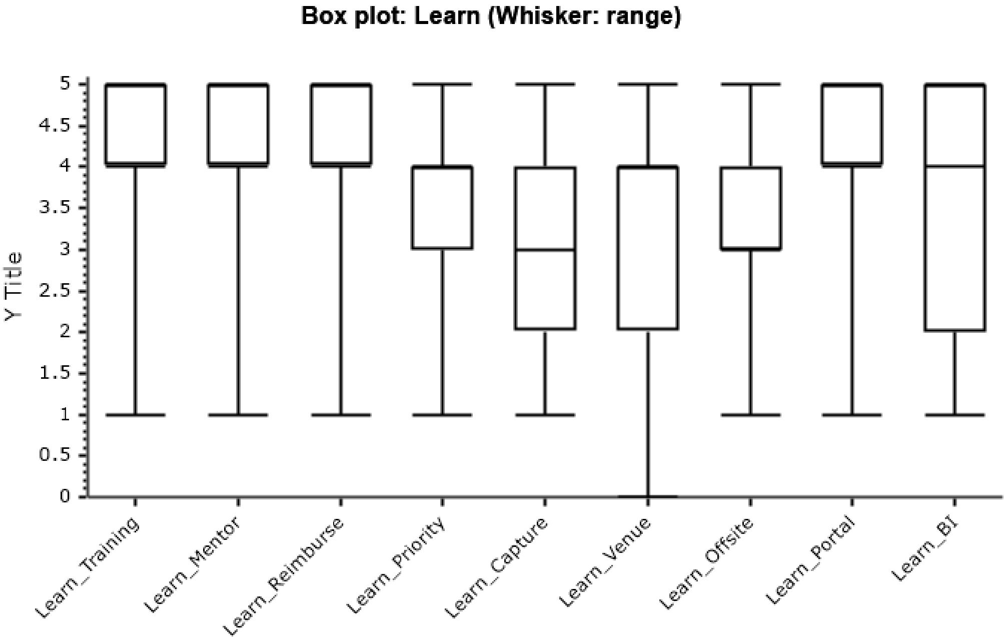


Fig. A3.45: Box plot of individual answers in the Learn category

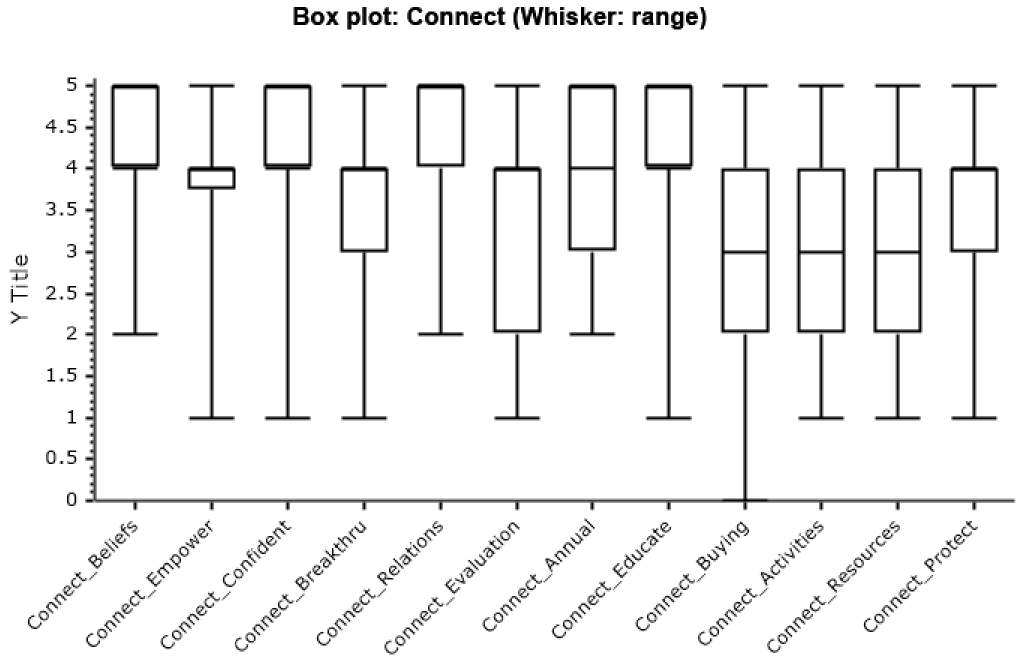


Fig. A3.46: Box plot of individual answers in the Connect category

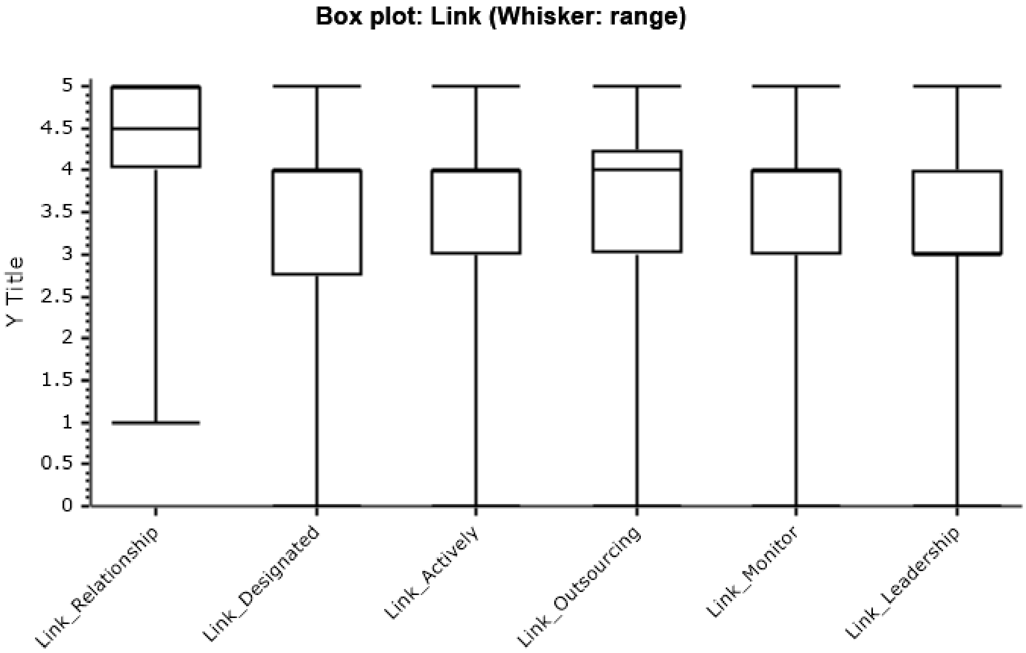


Fig. A3.47: Box plot of individual answers in the Link category

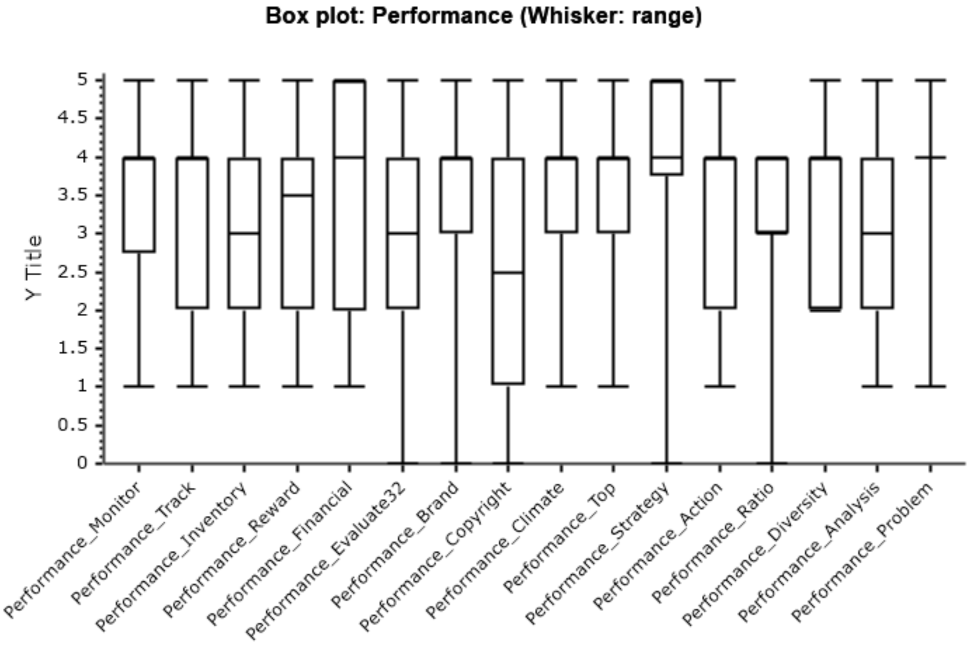


Fig. A3.48: Box plot of individual answers in the Performance category

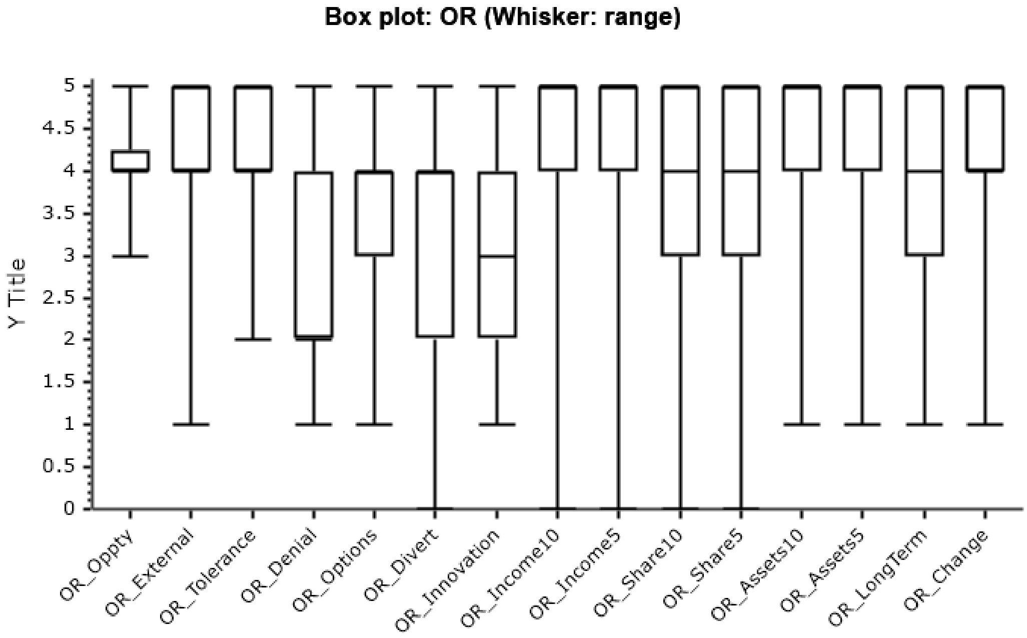
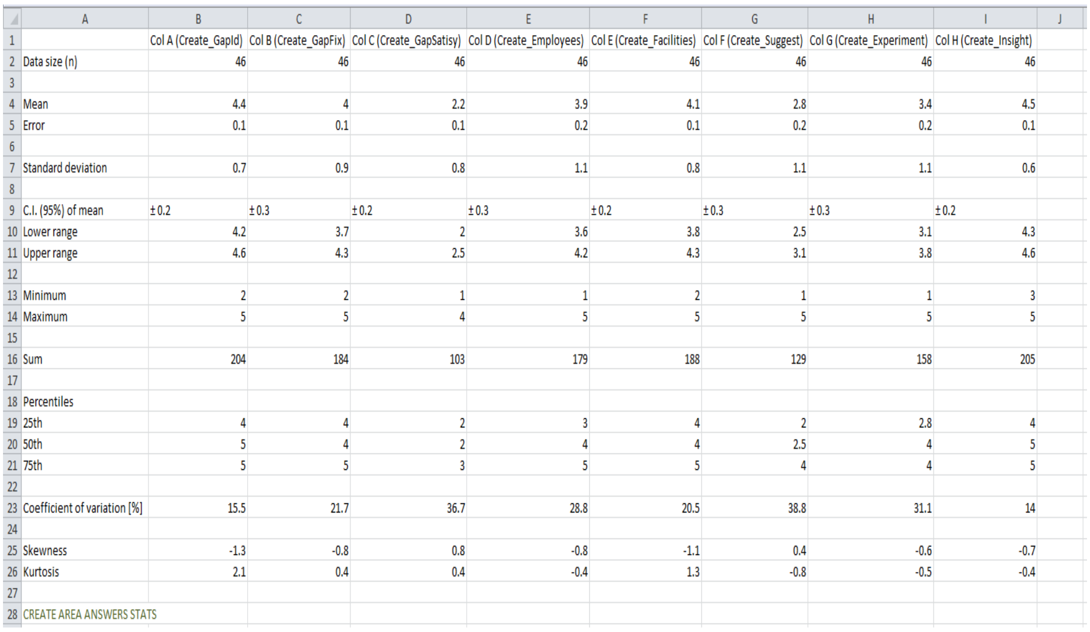
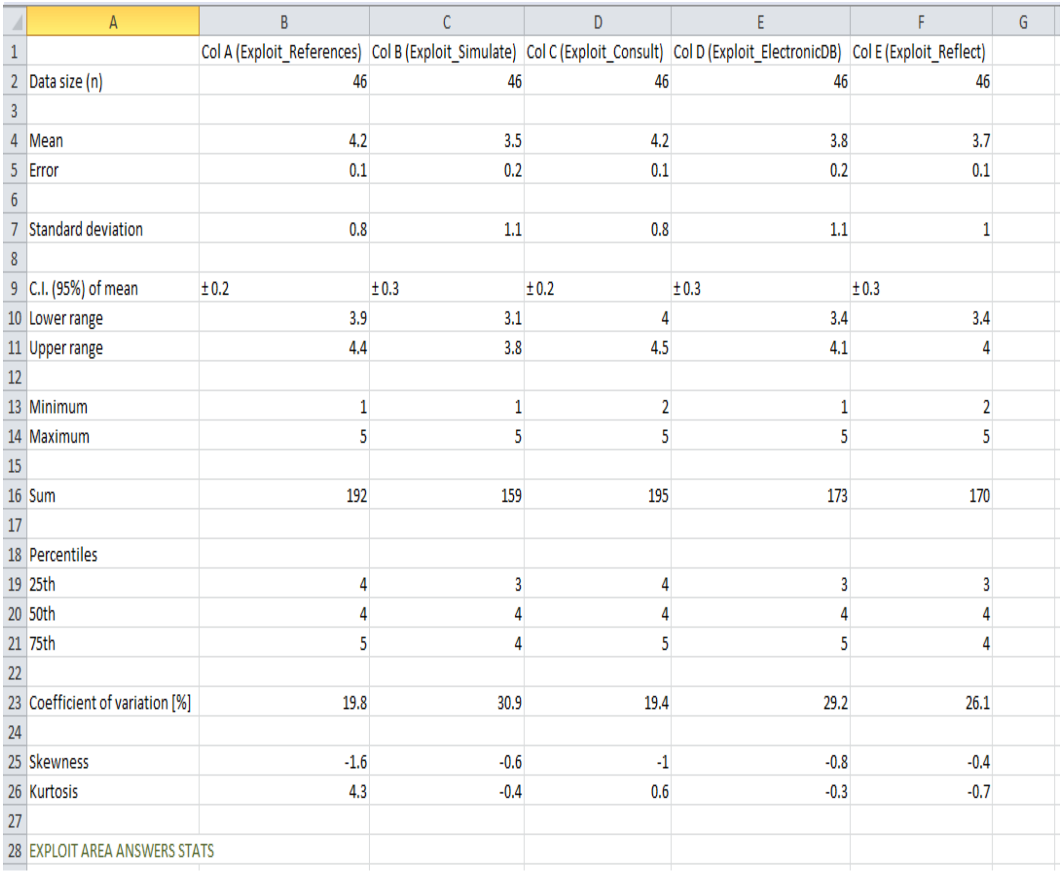


Fig. A3.49: Box plot of individual answers in the OR category



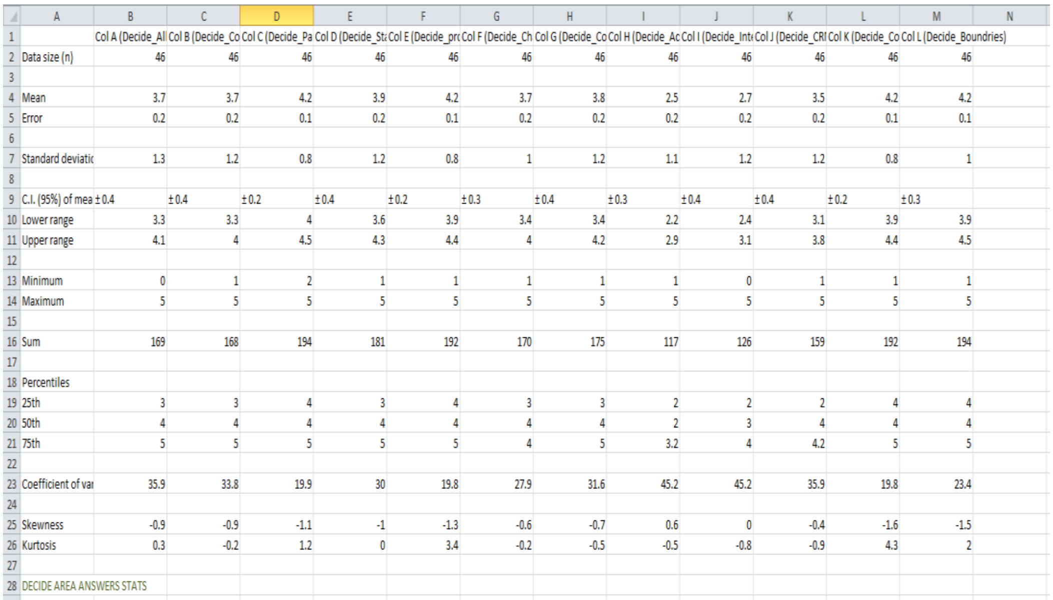
	A	B	C	D	E	F	G	H	I	J
1		Col A (Create_GapId)	Col B (Create_GapFix)	Col C (Create_GapSatisy)	Col D (Create_Employees)	Col E (Create_Facilities)	Col F (Create_Suggest)	Col G (Create_Experiment)	Col H (Create_Insight)	
2	Data size (n)	46	46	46	46	46	46	46	46	
3										
4	Mean	4.4	4	2.2	3.9	4.1	2.8	3.4	4.5	
5	Error	0.1	0.1	0.1	0.2	0.1	0.2	0.2	0.1	
6										
7	Standard deviation	0.7	0.9	0.8	1.1	0.8	1.1	1.1	0.6	
8										
9	C.I. (95%) of mean	±0.2	±0.3	±0.2	±0.3	±0.2	±0.3	±0.3	±0.2	
10	Lower range	4.2	3.7	2	3.6	3.8	2.5	3.1	4.3	
11	Upper range	4.6	4.3	2.5	4.2	4.3	3.1	3.8	4.6	
12										
13	Minimum	2	2	1	1	2	1	1	3	
14	Maximum	5	5	4	5	5	5	5	5	
15										
16	Sum	204	184	103	179	188	129	158	205	
17										
18	Percentiles									
19	25th	4	4	2	3	4	2	2.8	4	
20	50th	5	4	2	4	4	2.5	4	5	
21	75th	5	5	3	5	5	4	4	5	
22										
23	Coefficient of variation [%]	15.5	21.7	36.7	28.8	20.5	38.8	31.1	14	
24										
25	Skewness	-1.3	-0.8	0.8	-0.8	-1.1	0.4	-0.6	-0.7	
26	Kurtosis	2.1	0.4	0.4	-0.4	1.3	-0.8	-0.5	-0.4	
27										
28	CREATE AREA ANSWERS STATS									

Fig. A3.50: Statistics of the answers in the Create category



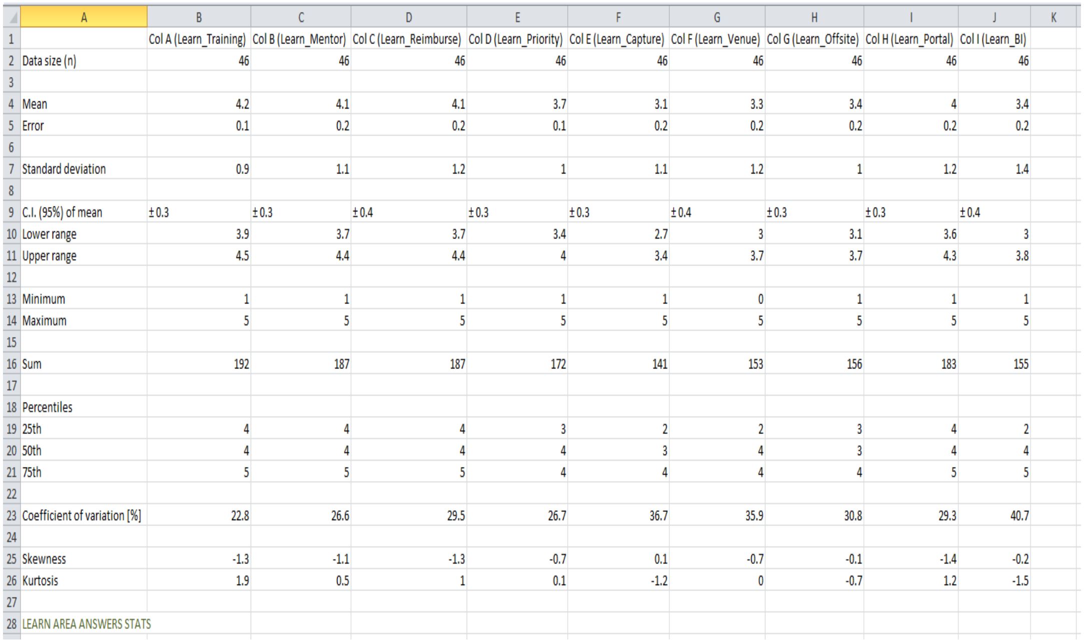
	A	B	C	D	E	F	G
1		Col A (Exploit_References)	Col B (Exploit_Simulate)	Col C (Exploit_Consult)	Col D (Exploit_ElectronicDB)	Col E (Exploit_Reflect)	
2	Data size (n)	46	46	46	46	46	
3							
4	Mean	4.2	3.5	4.2	3.8	3.7	
5	Error	0.1	0.2	0.1	0.2	0.1	
6							
7	Standard deviation	0.8	1.1	0.8	1.1	1	
8							
9	C.I. (95%) of mean	±0.2	±0.3	±0.2	±0.3	±0.3	
10	Lower range	3.9	3.1	4	3.4	3.4	
11	Upper range	4.4	3.8	4.5	4.1	4	
12							
13	Minimum	1	1	2	1	2	
14	Maximum	5	5	5	5	5	
15							
16	Sum	192	159	195	173	170	
17							
18	Percentiles						
19	25th	4	3	4	3	3	
20	50th	4	4	4	4	4	
21	75th	5	4	5	5	4	
22							
23	Coefficient of variation [%]	19.8	30.9	19.4	29.2	26.1	
24							
25	Skewness	-1.6	-0.6	-1	-0.8	-0.4	
26	Kurtosis	4.3	-0.4	0.6	-0.3	-0.7	
27							
28	EXPLOIT AREA ANSWERS STATS						

Fig. A3.51: Statistics of the answers in the Exploit category



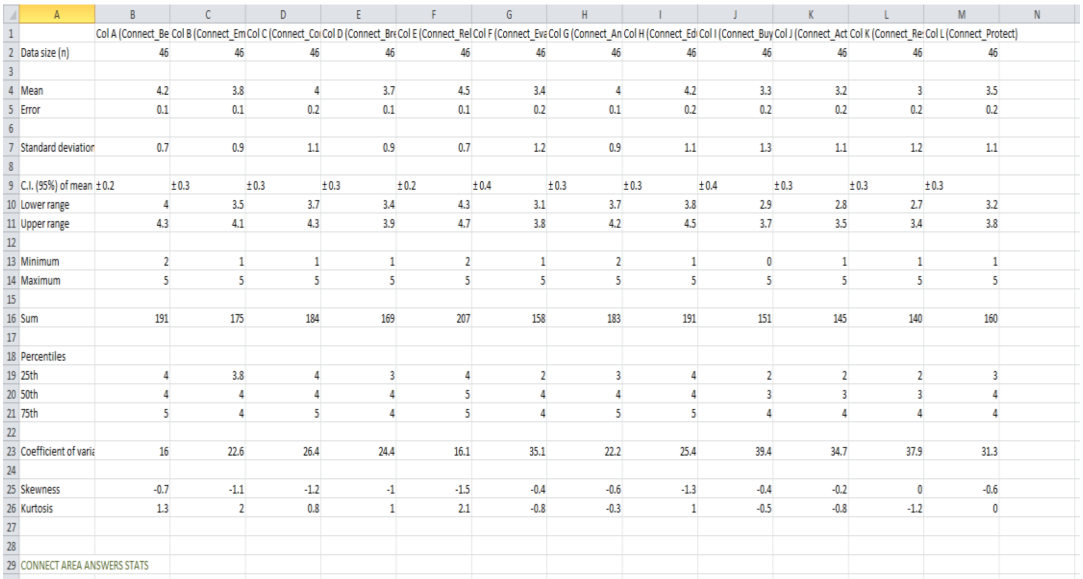
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		Col A (Decide_Al	Col B (Decide_Co	Col C (Decide_Pa	Col D (Decide_St	Col E (Decide_pr	Col F (Decide_Ch	Col G (Decide_Co	Col H (Decide_Ac	Col I (Decide_Int	Col J (Decide_CRI	Col K (Decide_Co	Col L (Decide_Boundries)	
2	Data size (n)	46	46	46	46	46	46	46	46	46	46	46	46	
3														
4	Mean	3.7	3.7	4.2	3.9	4.2	3.7	3.8	2.5	2.7	3.5	4.2	4.2	
5	Error	0.2	0.2	0.1	0.2	0.1	0.2	0.2	0.2	0.2	0.2	0.1	0.1	
6														
7	Standard deviatic	1.3	1.2	0.8	1.2	0.8	1	1.2	1.1	1.2	1.2	0.8	1	
8														
9	C.I. (95%) of mea	±0.4	±0.4	±0.2	±0.4	±0.2	±0.3	±0.4	±0.3	±0.4	±0.4	±0.2	±0.3	
10	Lower range	3.3	3.3	4	3.6	3.9	3.4	3.4	2.2	2.4	3.1	3.9	3.9	
11	Upper range	4.1	4	4.5	4.3	4.4	4	4.2	2.9	3.1	3.8	4.4	4.5	
12														
13	Minimum	0	1	2	1	1	1	1	1	0	1	1	1	
14	Maximum	5	5	5	5	5	5	5	5	5	5	5	5	
15														
16	Sum	169	168	194	181	192	170	175	117	126	159	192	194	
17														
18	Percentiles													
19	25th	3	3	4	3	4	3	3	2	2	2	4	4	
20	50th	4	4	4	4	4	4	4	2	3	4	4	4	
21	75th	5	5	5	5	5	4	5	3.2	4	4.2	5	5	
22														
23	Coefficient of var	35.9	33.8	19.9	30	19.8	27.9	31.6	45.2	45.2	35.9	19.8	23.4	
24														
25	Skewness	-0.9	-0.9	-1.1	-1	-1.3	-0.6	-0.7	0.6	0	-0.4	-1.6	-1.5	
26	Kurtosis	0.3	-0.2	1.2	0	3.4	-0.2	-0.5	-0.5	-0.8	-0.9	4.3	2	
27														
28	DECIDE AREA ANSWERS STATS													

Fig. A3.52: Statistics of the answers in the Decide category



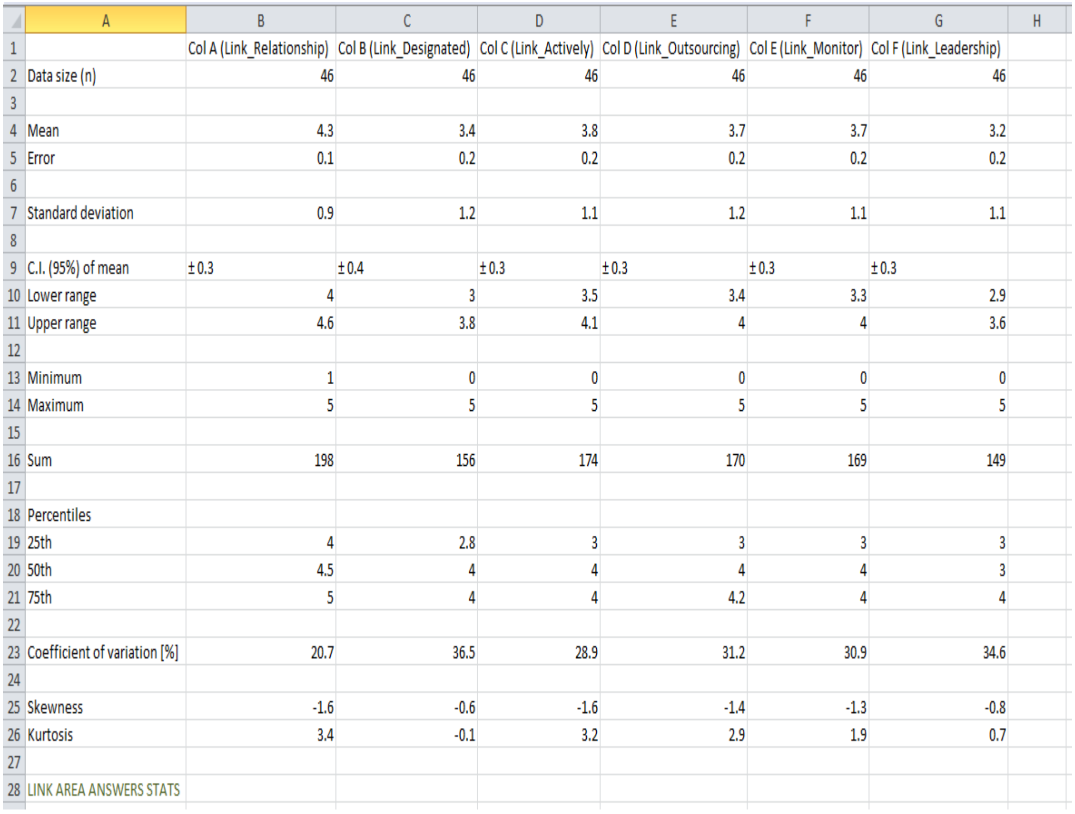
	A	B	C	D	E	F	G	H	I	J	K
1		Col A (Learn_Training)	Col B (Learn_Mentor)	Col C (Learn_Reimburse)	Col D (Learn_Priority)	Col E (Learn_Capture)	Col F (Learn_Venue)	Col G (Learn_Offsite)	Col H (Learn_Portal)	Col I (Learn_BI)	
2	Data size (n)	46	46	46	46	46	46	46	46	46	
3											
4	Mean	4.2	4.1	4.1	3.7	3.1	3.3	3.4	4	3.4	
5	Error	0.1	0.2	0.2	0.1	0.2	0.2	0.2	0.2	0.2	
6											
7	Standard deviation	0.9	1.1	1.2	1	1.1	1.2	1	1.2	1.4	
8											
9	C.I. (95%) of mean	±0.3	±0.3	±0.4	±0.3	±0.3	±0.4	±0.3	±0.3	±0.4	
10	Lower range	3.9	3.7	3.7	3.4	2.7	3	3.1	3.6	3	
11	Upper range	4.5	4.4	4.4	4	3.4	3.7	3.7	4.3	3.8	
12											
13	Minimum	1	1	1	1	1	0	1	1	1	
14	Maximum	5	5	5	5	5	5	5	5	5	
15											
16	Sum	192	187	187	172	141	153	156	183	155	
17											
18	Percentiles										
19	25th	4	4	4	3	2	2	3	4	2	
20	50th	4	4	4	4	3	4	3	4	4	
21	75th	5	5	5	4	4	4	4	5	5	
22											
23	Coefficient of variation [%]	22.8	26.6	29.5	26.7	36.7	35.9	30.8	29.3	40.7	
24											
25	Skewness	-1.3	-1.1	-1.3	-0.7	0.1	-0.7	-0.1	-1.4	-0.2	
26	Kurtosis	1.9	0.5	1	0.1	-1.2	0	-0.7	1.2	-1.5	
27											
28	LEARN AREA ANSWERS STATS										

Fig. A3.53: Statistics of the answers in the Learn category



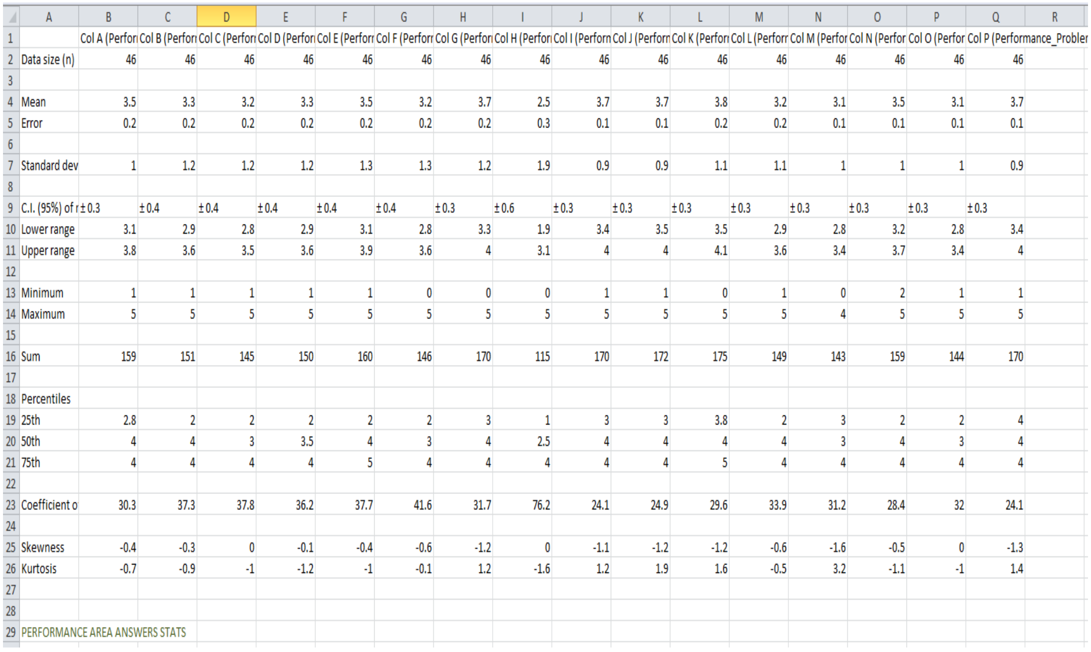
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		Col A (Connect_Be	Col B (Connect_Em	Col C (Connect_Co	Col D (Connect_Bre	Col E (Connect_Rel	Col F (Connect_Eva	Col G (Connect_An	Col H (Connect_Ed	Col I (Connect_Buy	Col J (Connect_Act	Col K (Connect_Re	Col L (Connect_Protect)	
2	Data size (n)	46	46	46	46	46	46	46	46	46	46	46	46	
3														
4	Mean	4.2	3.8	4	3.7	4.5	3.4	4	4.2	3.3	3.2	3	3.5	
5	Error	0.1	0.1	0.2	0.1	0.1	0.2	0.1	0.2	0.2	0.2	0.2	0.2	
6														
7	Standard deviation	0.7	0.9	1.1	0.9	0.7	1.2	0.9	1.1	1.3	1.1	1.2	1.1	
8														
9	C.I. (95%) of mean	±0.2	±0.3	±0.3	±0.3	±0.2	±0.4	±0.3	±0.3	±0.4	±0.3	±0.3	±0.3	
10	Lower range	4	3.5	3.7	3.4	4.3	3.1	3.7	3.8	2.9	2.8	2.7	3.2	
11	Upper range	4.3	4.1	4.3	3.9	4.7	3.8	4.2	4.5	3.7	3.5	3.4	3.8	
12														
13	Minimum	2	1	1	1	2	1	2	1	0	1	1	1	
14	Maximum	5	5	5	5	5	5	5	5	5	5	5	5	
15														
16	Sum	191	175	184	169	207	158	183	191	151	145	140	160	
17														
18	Percentiles													
19	25th	4	3.8	4	3	4	2	3	4	2	2	2	3	
20	50th	4	4	4	4	5	4	4	4	3	3	3	4	
21	75th	5	4	5	4	5	4	5	5	4	4	4	4	
22														
23	Coefficient of varia	16	22.6	26.4	24.4	16.1	35.1	22.2	25.4	39.4	34.7	37.9	31.3	
24														
25	Skewness	-0.7	-1.1	-1.2	-1	-1.5	-0.4	-0.6	-1.3	-0.4	-0.2	0	-0.6	
26	Kurtosis	1.3	2	0.8	1	2.1	-0.8	-0.3	1	-0.5	-0.8	-1.2	0	
27														
28														
29	CONNECT AREA ANSWERS STATS													

Fig. A3.54: Statistics of the answers in the Connect category



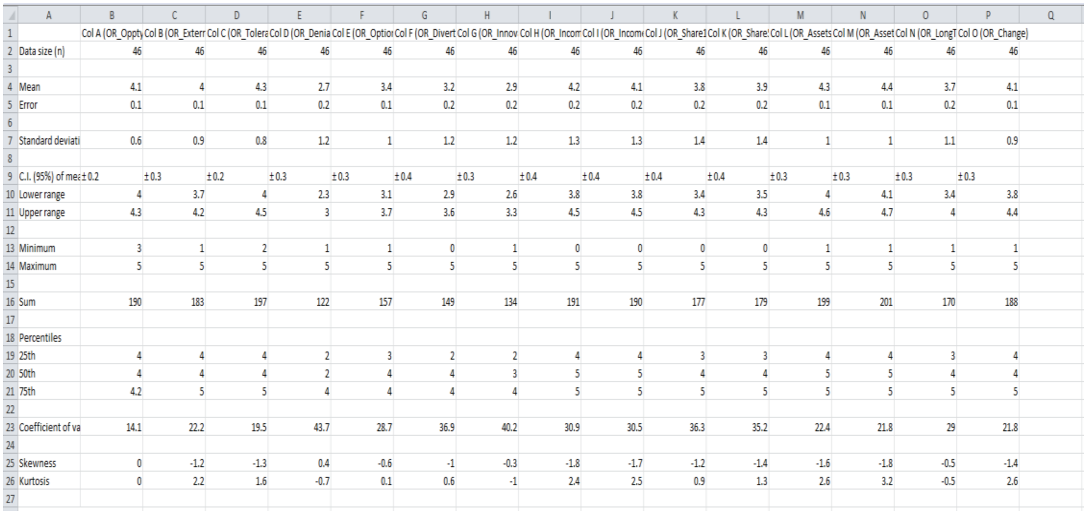
	A	B	C	D	E	F	G	H
1		Col A (Link_Relationship)	Col B (Link_Designated)	Col C (Link_Actively)	Col D (Link_Outsourcing)	Col E (Link_Monitor)	Col F (Link_Leadership)	
2	Data size (n)	46	46	46	46	46	46	
3								
4	Mean	4.3	3.4	3.8	3.7	3.7	3.2	
5	Error	0.1	0.2	0.2	0.2	0.2	0.2	
6								
7	Standard deviation	0.9	1.2	1.1	1.2	1.1	1.1	
8								
9	C.I. (95%) of mean	± 0.3	± 0.4	± 0.3	± 0.3	± 0.3	± 0.3	
10	Lower range	4	3	3.5	3.4	3.3	2.9	
11	Upper range	4.6	3.8	4.1	4	4	3.6	
12								
13	Minimum	1	0	0	0	0	0	
14	Maximum	5	5	5	5	5	5	
15								
16	Sum	198	156	174	170	169	149	
17								
18	Percentiles							
19	25th	4	2.8	3	3	3	3	
20	50th	4.5	4	4	4	4	3	
21	75th	5	4	4	4.2	4	4	
22								
23	Coefficient of variation [%]	20.7	36.5	28.9	31.2	30.9	34.6	
24								
25	Skewness	-1.6	-0.6	-1.6	-1.4	-1.3	-0.8	
26	Kurtosis	3.4	-0.1	3.2	2.9	1.9	0.7	
27								
28	LINK AREA ANSWERS STATS							

Fig. A3.55: Statistics of the answers in the Link category



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1		Col A (Perfor	Col B (Perfor	Col C (Perfor	Col D (Perfor	Col E (Perforr	Col F (Perforr	Col G (Perfor	Col H (Perfor	Col I (Perforn	Col J (Perforn	Col K (Perfor	Col L (Perforr	Col M (Perfor	Col N (Perfor	Col O (Perfor	Col P (Performance_Probler	
2	Data size (n)	46	46	46	46	46	46	46	46	46	46	46	46	46	46	46	46	
3																		
4	Mean	3.5	3.3	3.2	3.3	3.5	3.2	3.7	2.5	3.7	3.7	3.8	3.2	3.1	3.5	3.1	3.7	
5	Error	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.1	0.1	0.2	0.2	0.1	0.1	0.1	0.1	
6																		
7	Standard dev	1	1.2	1.2	1.2	1.3	1.3	1.2	1.9	0.9	0.9	1.1	1.1	1	1	1	0.9	
8																		
9	C.I. (95%) of r	± 0.3	± 0.4	± 0.4	± 0.4	± 0.4	± 0.4	± 0.3	± 0.6	± 0.3	± 0.3	± 0.3	± 0.3	± 0.3	± 0.3	± 0.3	± 0.3	
10	Lower range	3.1	2.9	2.8	2.9	3.1	2.8	3.3	1.9	3.4	3.5	3.5	2.9	2.8	3.2	2.8	3.4	
11	Upper range	3.8	3.6	3.5	3.6	3.9	3.6	4	3.1	4	4	4.1	3.6	3.4	3.7	3.4	4	
12																		
13	Minimum	1	1	1	1	1	0	0	0	1	1	0	1	0	2	1	1	
14	Maximum	5	5	5	5	5	5	5	5	5	5	5	5	4	5	5	5	
15																		
16	Sum	159	151	145	150	160	146	170	115	170	172	175	149	143	159	144	170	
17																		
18	Percentiles																	
19	25th	2.8	2	2	2	2	2	3	1	3	3	3.8	2	3	2	2	4	
20	50th	4	4	3	3.5	4	3	4	2.5	4	4	4	4	3	4	3	4	
21	75th	4	4	4	4	5	4	4	4	4	4	5	4	4	4	4	4	
22																		
23	Coefficient o	30.3	37.3	37.8	36.2	37.7	41.6	31.7	76.2	24.1	24.9	29.6	33.9	31.2	28.4	32	24.1	
24																		
25	Skewness	-0.4	-0.3	0	-0.1	-0.4	-0.6	-1.2	0	-1.1	-1.2	-1.2	-0.6	-1.6	-0.5	0	-1.3	
26	Kurtosis	-0.7	-0.9	-1	-1.2	-1	-0.1	1.2	-1.6	1.2	1.9	1.6	-0.5	3.2	-1.1	-1	1.4	
27																		
28																		
29	PERFORMANCE AREA ANSWERS STATS																	

Fig. A3.56: Statistics of the answers in the Performance category

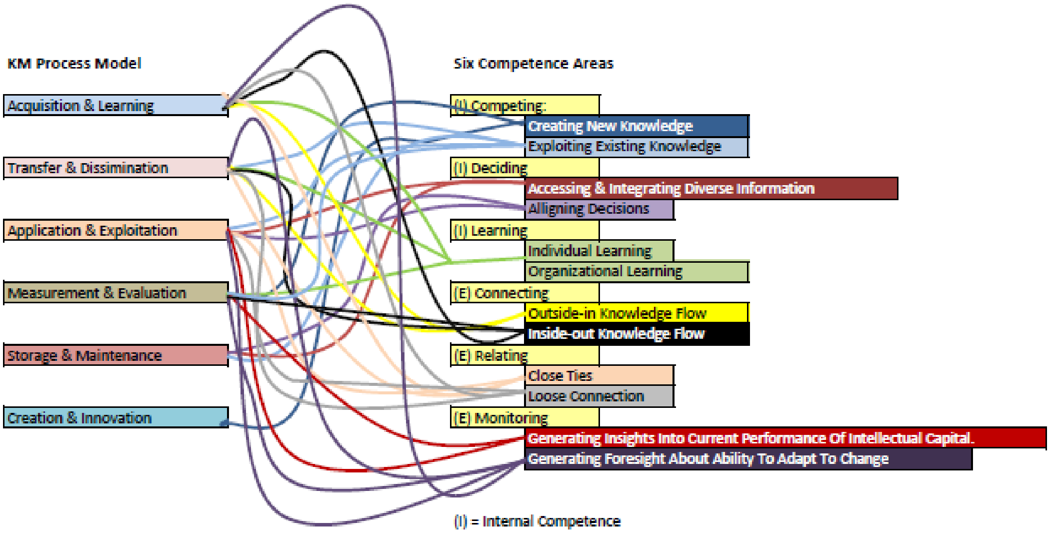


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1		Col A (OR_Oppty	Col B (OR_Exterr	Col C (OR_Tolera	Col D (OR_Denia	Col E (OR_Optior	Col F (OR_Divert	Col G (OR_Innov	Col H (OR_Incom	Col I (OR_Income	Col J (OR_Share1	Col K (OR_Share!	Col L (OR_Assets	Col M (OR_Asset	Col N (OR_LongT	Col O (OR_Change)	
2	Data size (n)	46	46	46	46	46	46	46	46	46	46	46	46	46	46	46	
3																	
4	Mean	4.1	4	4.3	2.7	3.4	3.2	2.9	4.2	4.1	3.8	3.9	4.3	4.4	3.7	4.1	
5	Error	0.1	0.1	0.1	0.2	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.1	0.1	0.2	0.1	
6																	
7	Standard deviati	0.6	0.9	0.8	1.2	1	1.2	1.2	1.3	1.3	1.4	1.4	1	1	1.1	0.9	
8																	
9	C.I. (95%) of mea	± 0.2	± 0.3	± 0.2	± 0.3	± 0.3	± 0.4	± 0.3	± 0.4	± 0.4	± 0.4	± 0.4	± 0.3	± 0.3	± 0.3	± 0.3	
10	Lower range	4	3.7	4	2.3	3.1	2.9	2.6	3.8	3.8	3.4	3.5	4	4.1	3.4	3.8	
11	Upper range	4.3	4.2	4.5	3	3.7	3.6	3.3	4.5	4.5	4.3	4.3	4.6	4.7	4	4.4	
12																	
13	Minimum	3	1	2	1	1	0	1	0	0	0	0	1	1	1	1	
14	Maximum	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	
15																	
16	Sum	190	183	197	122	157	149	134	191	190	177	179	199	201	170	188	
17																	
18	Percentiles																
19	25th	4	4	4	2	3	2	2	4	4	3	3	4	4	3	4	
20	50th	4	4	4	2	4	4	3	5	5	4	4	5	5	4	4	
21	75th	4.2	5	5	4	4	4	4	5	5	5	5	5	5	5	5	
22																	
23	Coefficient of va	14.1	22.2	19.5	43.7	28.7	36.9	40.2	30.9	30.5	36.3	35.2	22.4	21.8	29	21.8	
24																	
25	Skewness	0	-1.2	-1.3	0.4	-0.6	-1	-0.3	-1.8	-1.7	-1.2	-1.4	-1.6	-1.8	-0.5	-1.4	
26	Kurtosis	0	2.2	1.6	-0.7	0.1	0.6	-1	2.4	2.5	0.9	1.3	2.6	3.2	-0.5	2.6	
27																	

Fig. A3.57: Statistics of the answers in the Organizational Resilience category

APPENDIX IV: KM processes and Competence Areas Mapping

The Mapping of Most Common Influences of KM Processes on the Six Competence Areas and Vice Versa



(I) = Internal Competence
(E) = External Competence
Learning - Combined Categories Due To Extensive Similarities

Examples of functional/strategic relationship and outcome similarities between the concepts:

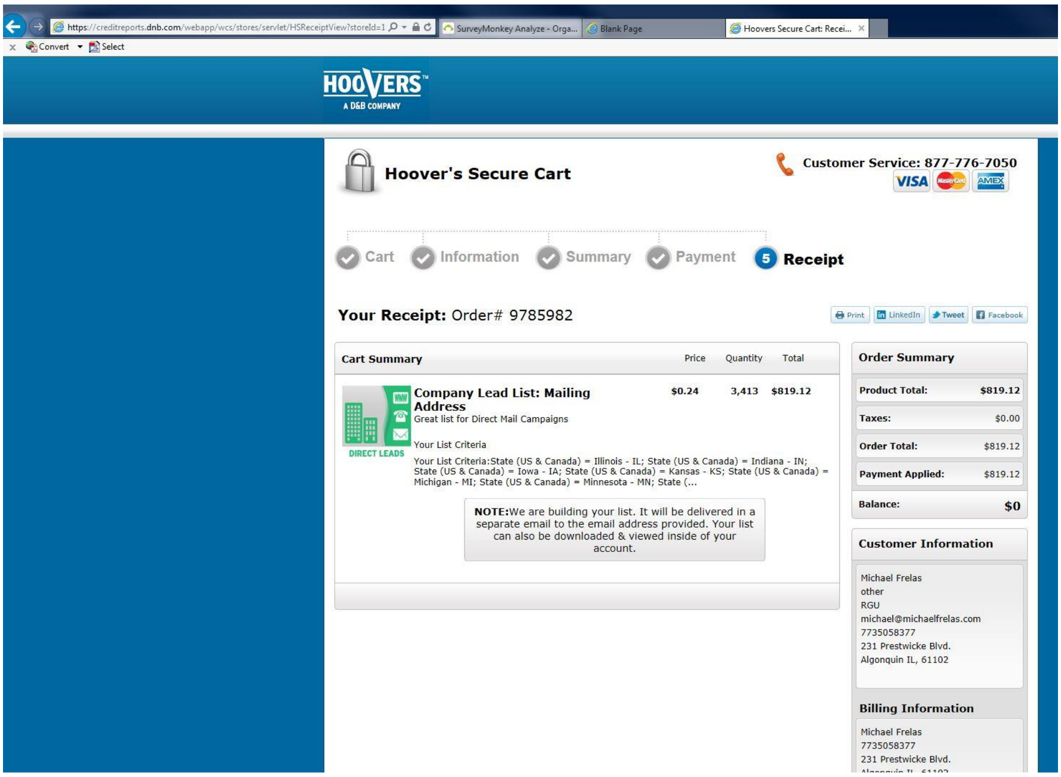
KM Processes: Six Competence Area Processes

Acquisition & Learning
Learning
Individual Learning
Formal lectures, on-the-job learning, reflection, mentoring, coaching, etc.
Organizational Learning
Somewhat similar to the individual learning with greater emphasis given to collective learning (involving connecting) through dialogue, stories, metaphores, CoPs.
Connecting
Outside-In Knowledge Flow
Environmental scanning, analyzing scans and identifying opportunities and threats.
Inside-Out Knowledge Flow
Learning from the result of releasing information to the outside environment.
Relating
Close Ties
Allows for learning/sharing of experiances among networked members while protecting against the competition. (Exchange of tacit and explicit knowledge.)
Loose Connections
Similar to the 'close ties' functions but emphasis on codified, easily diffusable knowledge.
Possible to learn through exteernal consultants, outsourcing.

Transfer & Dissimination
Competing
Exploiting Existing Knowledge
Provision of mechanisms allowing for the current knowledge to flow, in timely manner.
Learning
Individual Learning
Enable existing knowledge to flow for the purpose of learning.
Learning through interaction with others, reflection.
Organizational Learning
Dissimination of knowledge through Intranet, email, etc.
Knowledge databases. Lessosns learned.
Connecting
Outside-In Knowledge Flow
Knoweldge needed for environmental scanning needs to be transferred to the organization and, perhaps, propegated further within the company.
Inside-Out Knowledge Flow
Release of information expected to benefit the organziation upon release

	Relating		
		Close Ties	
			Trust between closely connected business partners provides environment for knowledge and idea exchange.
		Loose Connections	
			Exchange/release of, mainly, explicit knowledge transferred with limits due to limited trust.
	Monitoring		
		Generating Foresight About Ability To Adapt To Change	
			Ability to transfer outomes of monitoring so that it can be acted upon. Ability to share the foresights with the stakeholders.
Application & Exploitation	Competing		
		Exploiting Existing Knowledge	
			Selling patents. Utilizing existing knowledge in improvement of in-house operations. Knowledge repositories, databases, expert systems.
	Deciding		
		Accessing & Integrating Diverse Information	
			Use accumulated in-house knowledge for decision making, paying attention to diversity of information and knowledge.
	Deciding		
		Alliging Decision	
			Similar to the function described above but making sure the activities are harmonized across all levels of organization.
	Relating		
		Close Ties	
			Close ties perform a secondary function of application and exploration: they provide a channel for cooperation and relationship building.
		Loose Connections	
			Connections can still serve, although not in as large extent as it is the case with close ties, as a medium in cooperation, project participation, etc.
	Monitoring		
		Generating Insights Into Current Performance Of Intellectual Capital.	
			Evaluation of utilization and value creation of current knowledge assets.
		Generating Foresight About Ability To Adapt To Change	
			Monitoring the external environment for the 'first mover' advantage. Anticipating market/competitiors actions and movements.
Measurement & Evaluation	Competing		
		Creating New Knowledge	
			Assesment of the potential value of new knowledge.
	Learning		
		Individual Learning	
			Post course quetionaires, follow-up inquires.
		Organizational Learning	
			Performance measurement, seaking measurable improvements as a result of organziational learning.
	Connecting		
		Inside-Out Knowledge Flow	
			Evaluation and measurement of released knowledge on the market, competitiors, etc.
	Monitoring		
		Generating Insights Into Current Performance Of Intellectual Capital.	
			Evaluation of the use of exisintg in-house knowledge.
		Generating Foresight About Ability To Adapt To Change	
			Measure external enviornment to detect signal/s for the need to change. Measure usefullnes of competitor and market analysis. Measure the internal capabilities with relation to possible direction of strategy change.
Storage & Maintenance	Competing		
		Expoiting Existing Knowledge	
			Need to store, maintain and retrieve knowledge when needed, in a timely fashion.
	Deciding		
		Accessing & Integrating Diverse Information	
			Making it easy to access similar cases/circumstances and decision made in those cases. Recording new decisions and decision making steps and justifications.
		Alligning Decisions	
			Same as above but making sure this is harmonized across all levels of organziation.
Creation & Innovation	Competing		
		Creating New Knowledge	
			Generation of knowledge about markets, customers, suppliers, partners, etc.

APPENDIX V: Organization Mailing List



https://creditreports.dnb.com/webapp/wcs/stores/servlet/HSReceiptView?storeId=1
SurveyMonkey Analyze - Orga...
Blank Page
Hoovers Secure Cart: Recei...
Convert
Select
HOOVERS™
A D&B COMPANY
Cart
Information
Summary
Payment
5
Receipt
Print
LinkedIn
Tweet
Facebook
DIRECT LEADS

Hoover's Secure Cart

Customer Service: 877-776-7050

Your Receipt: Order# 9785982

Cart Summary	Price	Quantity	Total
Company Lead List: Mailing Address Great list for Direct Mail Campaigns Your List Criteria Your List Criteria:State (US & Canada) = Illinois - IL; State (US & Canada) = Indiana - IN; State (US & Canada) = Iowa - IA; State (US & Canada) = Kansas - KS; State (US & Canada) = Michigan - MI; State (US & Canada) = Minnesota - MN; State (...	**$0.24**	**3,413**	**$819.12**

NOTE: We are building your list. It will be delivered in a separate email to the email address provided. Your list can also be downloaded & viewed inside of your account.

Order Summary	
Product Total:	**$819.12**
Taxes:	$0.00
Order Total:	$819.12
Payment Applied:	$819.12
Balance:	**$0**

Customer Information

Michael Frelas
other
RGU
michael@michaelfrelas.com
7735058377
231 Prestwicke Blvd.
Algonquin IL, 61102

Billing Information

Michael Frelas
7735058377
231 Prestwicke Blvd.
Algonquin IL, 61102

A5.1: Business leads purchased from Dunn & Bradstreet's sister company: Hoover

APPENDIX VI: Company List Processing Steps

20140126: List, consisting of 3,413 records has been received from Dun & Bradstreet's sister company: Hoovers.com. The list was received in a commas delimited flat file (csv format).

List was cleaned up. Six duplicate records were deleted.

20140202: List, in the csv format was first loaded into Excel file. Then the Excel file was arranged, alphabetically (from A to Z) by the company's name.

20140205: Creating program, using Microsoft's Visual Basic for Applications that would select 1000 records out of the total of 3413 company names [representing 29% of the entire population]. Some of the titles of executives were changed from Cfo to CFO, from Cto to CTO, from Ceo to CEO.

20140205: Selected, randomly (using VBA's function: Int ((upperbound - lowerbound + 1) * Rnd + lowerbound) first 1000 records to send out. In the formula the upper bound was replaced by 3413 and lower bound by 1).

20140206: Five companies (out of the initial 1000) did not have contact person specified. The letter was addressed to the 'President' in these five cases. [All other information was present.]

20140208: 1000 envelopes (security type that do not allow 'see through') were stuffed with the letters and sealed.

20140210: 1000 envelopes were affixed with shipping label.

20140211: Cleared all test/trial responses to the survey – getting the website (at www.surveymonkey.com) ready to accept the 'real' input.

20140212: Purchased 1000 first-class stamps.

20140213: Sent 1000 letters – dropped off inside the post office @ 9:20am. The letters are to arrive at the destinations on Tue-Wed (Feb 18-19) due to Federal holiday on Monday, Feb 17. [There is standard 2-3 day delivery window for first-class mail for mail sent within continental US.]

20140301 – 20140309: Prepared 1,000 new letters: folded, prepared and sealed envelopes.

20140306 – One returned letter, from Landmark Bank in Columbia, MO, stating that the person that the letter was addressed to has retired.

20140311 – Purchased 1,000 stamps and applied them on envelopes.

Illustration of the sent out letters:

[redacted]
[redacted]

20140301: Sent 12 emails to local Chambers of Commerce asking for help with questionnaire completion.

20140305: Due to lack of any replies to emails sent on March 1, 12 letters were mailed to the original Chamber of Commerce contacted initially via email.

20140305: Contacted, via email, seven National Associations asking for help with questionnaire completion.

20140306 – Preparing the next mailing of 1,000 envelopes.

- Deleted 1,000 previously printed company names and contacts + 2 companies that were duplicates. Total removed record 3413 – 1002, leaving 2411 records.
- Manually deleted non-profit universities and, high schools, school districts and etc. [There will be a need to exclude replies from such organizations in response to the first mailing batch.] This action resulted in remaining 2221 records down from 2411.
- 1,000 new records, out of 2221, were randomly selected using previously described algorithm.
- 8 out of 1,000 selected records had no contact information. The blank space has been replaced by the words "Company President".
- Company name, contact name and address were all capitalized.

20140315: Wrote letter to the Landmark Bank asking for completion of the questionnaire. This is the organization that sent the letter informing about the retirement of the key executive to whom the first letter was addressed. (See note from 20140311).

20140317: 1,000 letters were mailed at 9:10am from the post office in Algonquin, IL.

20140422: Sent an invitation to complete the questionnaire to sixteen people known personally that are in senior positions. The following is the list of companies (with the company and executive names removed).

- Solution Director [Consulting Company]
- CIO [Large Insurance Company]
- Director, Enterprise Apps. [Semiconductor Company]
- Solution Architect [Major Airline]
- Sr. IT Mgr. [Large Insurance Company]
- VCEO [Analytics Company]
- VP, Supply Chain [Large Retailer]
- VP, Business Analysis [Large Insurance Company]
- Director, Pricing [Telecomm]
- Program Director [Television]
- IT Director [Large Recruiting Firm]
- Chief Functional Architect [Mid-size Software Firm]
- CFO [Mid-size Medical Software Company]
- CEO [Mid-size Consulting (Software) Company]
- Director of Applications [Small Software Firm]
- VP, Claims [Large Insurance Company]

The above people received the following email via LinkedIn messaging system:

Dear Executive,

I recognize that the demands on your time are enormous, so my appreciation for your participation in this academic research project cannot be overstated. I am truly grateful for your time, and I hope to provide something of value for your organization in return for approximately 15 minutes of your time. This questionnaire is a chance for you to state your opinions for the benefit of businesses based in Midwest as well as the benefits of society.

By completing this questionnaire, you will be contributing to research in the field. Your input is of great value to this work and is greatly appreciated. In return for your time devoted to answering this questionnaire you will be provided (free of charge) with a feedback on your organization’s responses vs. other participating companies. Please note, any responses you provide will be treated confidentially, and your anonymity will be preserved.

Questionnaire’s link: https://www.surveymonkey.com/s/MFrelas

Once again, I would like to thank you for your time.

Please do not hesitate to contact me or my research supervisor (Dr. Burnett s.burnett@rgu.ac.uk) if you have any questions or comments related to this research. If you feel that someone else at your organization should be responding to this questionnaire then please pass this questionnaire along – thank you!

Best regards,

Michael Frelas, Doctoral Candidate

m.frelas@rgu.ac.uk

Cell phone #: (773) 505-8377

20140504: Sending the above quoted letters (via snail mail) to people known in the past whose business cards have been located in personal business card collection. The letters were sent to:

- CIO [Law Firm]
- President [Kitchen & Baths Distributor]
- President [Steel Manufacturing & Construction]
- General Manager [Medical Device Manufacturer]
- VP of Purchasing [Boat Manufacturing]
- CFO [Manufacturer and Distributor of Collectibles]
- President [Precision Parts Manufacturer]

20140518: Changed the content of the letter to the executives: exchanged the '20 minutes' by '10 minutes'.

20140522: Preparing the last batch of 1,219 letters. 25 letters (in nearby area of IL were stuffed with business card, in addition to the letter).

20140524: During the application of printed labels additional 8 entries were discarded (these were public school districts). The total number of mail pieces sent from the purchased 3,413 addresses/names was: 3,211 [three thousand two hundred and eleven. In three batches: 1,000 + 1,000 + 1,211].

20140526: Completed stuffing 1,211 envelopes. (Also, all envelopes were affixed address labels and were sealed.)

20140530: Purchased 1,211 first class stamps.

20140531: Applied stamps on the envelopes and mailed all envelopes in Algonquin's post office around 10:30am. This brings the total of mailed pieces to: 3 211.

20141216: Posted to the Research Methods and Data Science the following post:

Title: Replication of survey data for data mining purposes in doctoral research - seeking suggestions, thank you.

Body: Hello. Thank you for reading and my apologies if my posting is not appropriate for the group. I am completing survey-based doctoral research that uses data mining (Naïve Bayes, Data Trees provided by Microsoft's SQL Server) as a tool for data analysis. I sent out 3,200 surveys but have only received 38 fully completed surveys back - perhaps due to the fact that each survey contains 84 questions. Because of my research needs to use data mining for survey reply analysis I have replicated each answer 10,000 times so the present count of 'replies' equals 380,000. While this number of records allows for interesting data mining, I am not sure how to take this further. Clearly, no general assumptions can be made about the entire population but what about validity of conclusions assuming the answers received apply to the larger sample or even to the entire population? Or, perhaps I should look at some other aspect for completion of my research (and this is year # 6 of studies so it would be nice to be able to complete set out goal ;-) Thank you for all comments/insights. Best wishes.

Awaiting replies.

20141217: Posted the following text to Research, Methodology, and Statistics in the Social Science.

Hello. Thank you for reading and my apologies if my posting is not appropriate for the group. I am completing survey-based doctoral research that uses data mining (Naïve Bayes, Data Trees provided by Microsoft's SQL Server) as a tool for data analysis. I sent out 3,200 surveys but have only received 38 fully completed surveys back - perhaps due to the fact that each survey contains 84 questions.

The surveys attempted to measure Organizational Resilience and they were sent to the CEOs of the mid-size companies located in the mid-west (USA). 3200 companies, according to my source and my definition of mid-size and mid-west, constitute the entire population. The names and addresses of CEOs/firms were purchased from the prime source: Dun & Bradstreet and were considered accurate. In return for the completion of the survey, CEOs were offered analysis of their answers as relation to all other answers + other small benefits.

The pilot survey was sent out prior to the mass mailing of 3200 surveys and it returned all positive feedback (8 pilot surveys returned out of 10 mailed out). The avg. time to complete the survey was around 10 minutes. Here is the most important part: In the survey questions were categorized into 8 categories, with 5-15 questions per category. Answers to the questions were assigned points and summed up for each category, per each reply. Then 7 out of 8 categories (summed up points for answers within each category) were used as input (independent variable) and the 8th category became the dependent variable (predict only). Because of very low return rate and data mining's need of large data quantities each reply was replicated 10 000 times bringing the number of records submitted to data mining algorithms to 380 000.

Finally, data mining algorithms (Naïve Bayes and Decision Tree) were used against 380 000 records to determine relationships between the 7 input and 1 predict categories. Having said all of this, I wonder how does the replication of replies by 10 000 affects the analysis of the correlation between 7 input and 1 output variable? Is all of the work completed thus far unusable? Any suggestions about the approach or ideas about salvaging this research will be greatly appreciated – thank you!

Awaiting replies.

Here is the complete posting, as of 12/21/14, from the group:

- Irma

  Irma Diaz-Martin

  Investigator at State of California

  Hi Michael,

  I also am in the final stages of my dissertation and definitely understand the frustration of unexpected road blocks!! In reading your post and reviewing your profile, it appears you definitely have the expertise for the development of valid surveys and are well versed in analytical systems. The first thing I noticed in your post was the huge sample population! Could it be possible to redefine your sample population so that it is smaller, therefore providing you a more viable sample while increasing the response rate? For example, could your sample population be "mid-size companies in the greater Chicago area" or even better yet, within a district or county, rather than the "Midwest?" I'm sure this would require you consulting with your dissertation committee, but it may be your sample population is too large and also difficult to reach since you are targeting CEOs.

  Also the size of your survey instrument may not have a bearing in your response rate. It may simply be the responsiveness of your sample population. I administered a 112 question survey successfully but received
  great guidance from my committee which resulted in a viable sample

population.

Just a thought. Wish you the best in the final lap of your dissertation.

- Unlike Like
- Reply privately
- Flag as inappropriate
- 2 days ago

- Brigid

  Brigid McDermott

  Research Methods Professional

  Hi Michael.
  I was drawn to your post because I sympathize with everyone whose graduate research throws them a curve ball.
  Certainly there is no statistical magic that will multiply the actual information you have in your 38 surveys.
  Is it possible to move your graduate research in the direction of methodology in your discipline? You mention various methods. Are the strengths and weaknesses of these methods understood for your discipline? Are therer better methods? You could try simulating data with various attributes and then testing these and other methods to answer these questions.
  Do you need to get the information you requested directly from the CEO's or could you scrape the company websites for the information you need?
  Then there is the question of survey response. How does one get a survey completed in an age of "survey monkey" where requests for completing surveys has almost become spam in one's inbox? Can you do an analysis of the difference in the CEO's who completed your survey and those that did not? Are there any lessons to be learned here? Can you test the lessons learned by reissuing your survey having implemented the changes from the "lessons learned" and see if you get an increased response rate?
  I hope you find a way to proceed with your graduate research.

  - Unlike Like
  - Reply privately
  - Flag as inappropriate
  - 2 days ago

- Michael Frelas

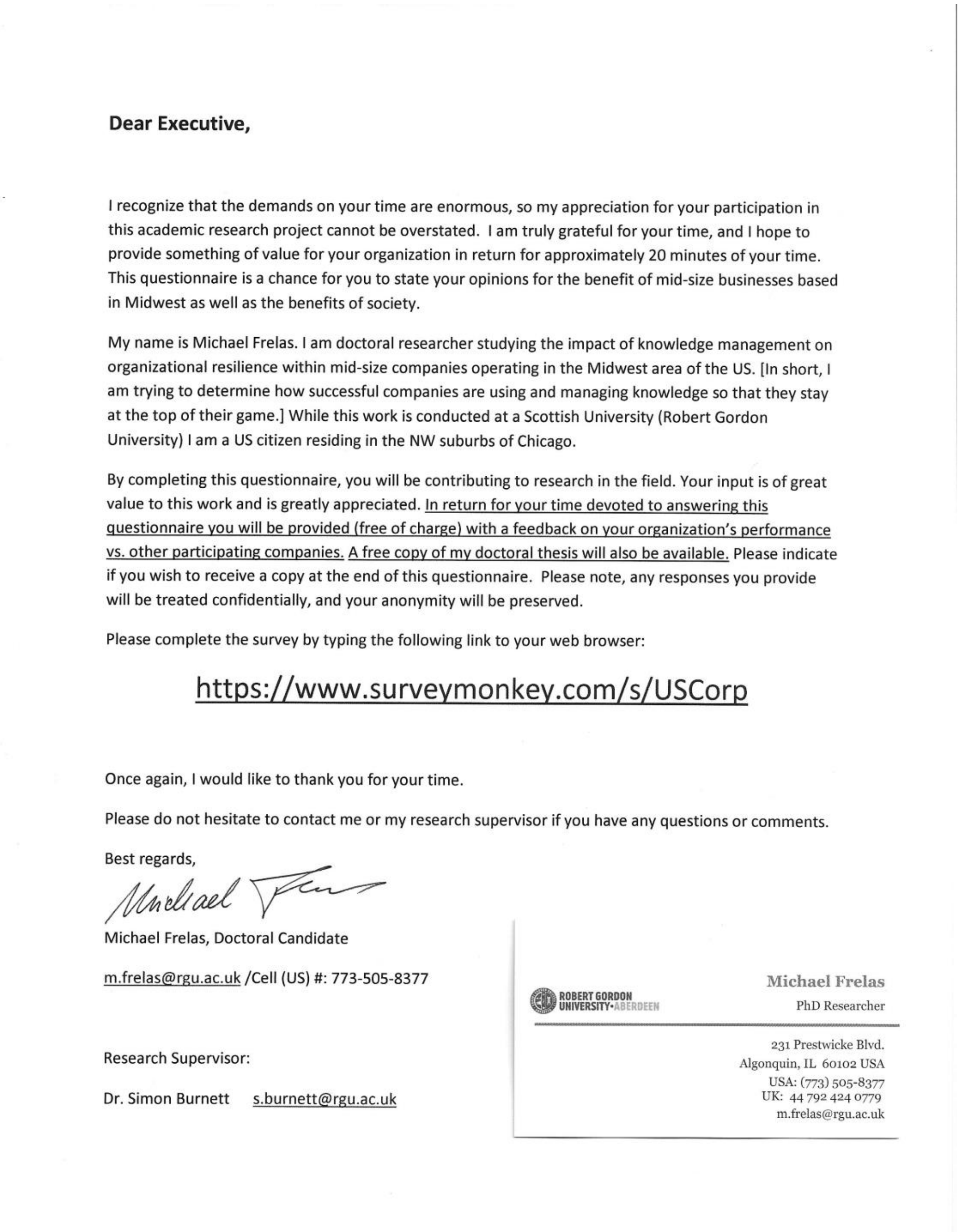
  Business Intelligence, Data Warehouse & Analytics Architect. PhD Researcher @ RGU, United Kingdom.

Top Contributor

Dear Irma and Bridig - I greatly appreciate your replies and your suggestions, thank you very much. Let me spend the next week or two on reviewing your suggestions. Again, thank you so much! Happy Holidays!

- Delete
- 2 days ago

APPENDIX VII: Letter to the Executives



Dear Executive,

I recognize that the demands on your time are enormous, so my appreciation for your participation in this academic research project cannot be overstated. I am truly grateful for your time, and I hope to provide something of value for your organization in return for approximately 20 minutes of your time. This questionnaire is a chance for you to state your opinions for the benefit of mid-size businesses based in Midwest as well as the benefits of society.

My name is Michael Frelas. I am doctoral researcher studying the impact of knowledge management on organizational resilience within mid-size companies operating in the Midwest area of the US. [In short, I am trying to determine how successful companies are using and managing knowledge so that they stay at the top of their game.] While this work is conducted at a Scottish University (Robert Gordon University) I am a US citizen residing in the NW suburbs of Chicago.

By completing this questionnaire, you will be contributing to research in the field. Your input is of great value to this work and is greatly appreciated. In return for your time devoted to answering this questionnaire you will be provided (free of charge) with a feedback on your organization's performance vs. other participating companies. A free copy of my doctoral thesis will also be available. Please indicate if you wish to receive a copy at the end of this questionnaire. Please note, any responses you provide will be treated confidentially, and your anonymity will be preserved.

Please complete the survey by typing the following link to your web browser:

https://www.surveymonkey.com/s/USCorp

Once again, I would like to thank you for your time.

Please do not hesitate to contact me or my research supervisor if you have any questions or comments.

Best regards,

Michael Frelas, Doctoral Candidate

m.frelas@rgu.ac.uk /Cell (US) #: 773-505-8377

Research Supervisor:

Dr. Simon Burnett s.burnett@rgu.ac.uk

ROBERT GORDON
UNIVERSITY·ABERDEEN

Michael Frelas
PhD Researcher

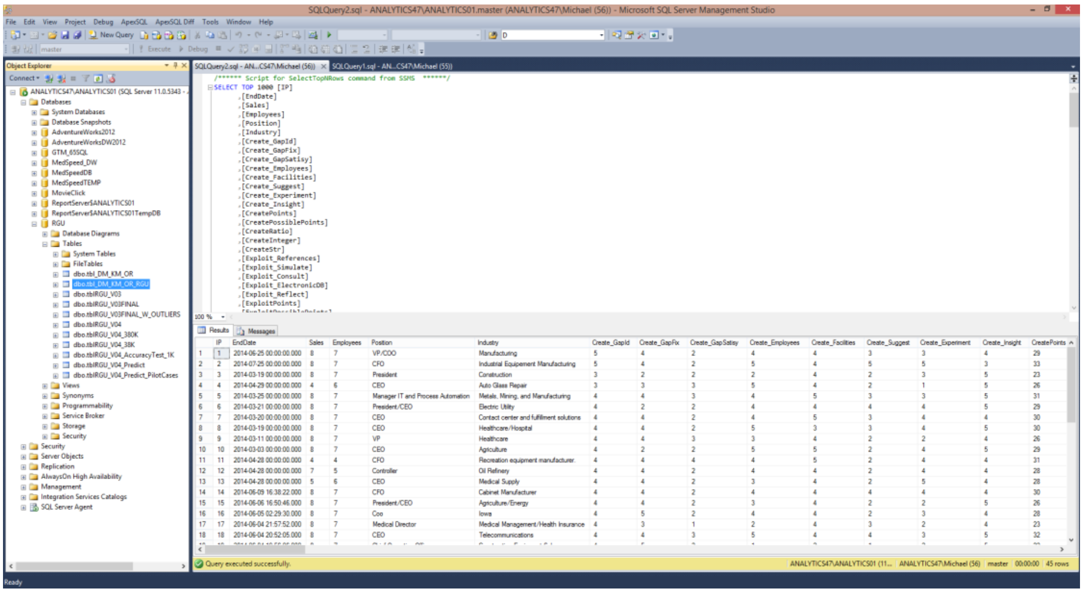
231 Prestwicke Blvd.
Algonquin, IL 60102 USA
USA: (773) 505-8377
UK: 44 792 424 0779
m.frelas@rgu.ac.uk

APPENDIX VIII: Data Mining Supporting Documentation

Data Mining

```
USE [RGU]
GO

/****** Object:  Table [dbo].[tbl_DM_KM_OR_RGU]    Script Date: 6/17/2016
6:38:58 AM ******/
SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[tbl_DM_KM_OR_RGU](
        [IP] [int] NOT NULL,
        [EndDate] [datetime] NULL,
        [Sales] [tinyint] NULL,
        [Employees] [tinyint] NULL,
        [Position] [nvarchar](255) NULL,
        [Industry] [nvarchar](255) NULL,
        [Create_GapId] [smallint] NULL,
        [Create_GapFix] [tinyint] NULL,
        [Create_GapSatisy] [tinyint] NULL,
        [Create_Employees] [tinyint] NULL,
        [Create_Facilities] [tinyint] NULL,
        [Create_Suggest] [tinyint] NULL,
        [Create_Experiment] [tinyint] NULL,
        [Create_Insight] [tinyint] NULL,
        [CreatePoints] [tinyint] NULL,
        [CreatePossiblePoints] [tinyint] NULL,
        [CreateRatio] [float] NULL,
        [CreateInteger] [tinyint] NULL,
        [CreateStr] [nvarchar](255) NULL,
        [Exploit_References] [tinyint] NULL,
        [Exploit_Simulate] [tinyint] NULL,
        [Exploit_Consult] [tinyint] NULL,
        [Exploit_ElectronicDB] [tinyint] NULL,
        [Exploit_Reflect] [tinyint] NULL,
        [ExploitPoints] [tinyint] NULL,
        [ExploitPossiblePoints] [tinyint] NULL,
        [ExploitRatio] [float] NULL,
        [ExploitInteger] [tinyint] NULL,
        [ExploitStr] [nvarchar](255) NULL,
        [Decide_Alliances] [tinyint] NULL,
        [Decide_CoOp] [tinyint] NULL,
        [Decide_Partnership] [tinyint] NULL,
        [Decide_Standards] [tinyint] NULL,
        [Decide_professional] [tinyint] NULL,
        [Decide_Chambers] [tinyint] NULL,
        [Decide_Communities] [tinyint] NULL,
        [Decide_Academic] [tinyint] NULL,
        [Decide_Intelligence] [tinyint] NULL,
        [Decide_CRM] [tinyint] NULL,
        [Decide_Condition] [tinyint] NULL,
        [Decide_Boundries] [tinyint] NULL,
        [DecidePoints] [tinyint] NULL,
        [DecidePossiblePoints] [tinyint] NULL,
        [DecideRatio] [float] NULL,
        [DecideInteger] [tinyint] NULL,
        [DecideStr] [nvarchar](255) NULL,
        [Learn_Training] [tinyint] NULL,
        [Learn_Mentor] [tinyint] NULL,
        [Learn_Reimburse] [tinyint] NULL,
        [Learn_Priority] [tinyint] NULL,
        [Learn_Capture] [tinyint] NULL,
        [Learn_Venue] [tinyint] NULL,
        [Learn_Offsite] [tinyint] NULL,
        [Learn_Portal] [tinyint] NULL,
        [Learn_BI] [tinyint] NULL,
        [LearnPoints] [tinyint] NULL,
        [LearnPossiblePoints] [tinyint] NULL,
        [LearnRatio] [float] NULL,
        [LearnInteger] [tinyint] NULL,
        [LearnStr] [nvarchar](255) NULL,
        [Connect_Beliefs] [tinyint] NULL,
        [Connect_Empower] [tinyint] NULL,
        [Connect_Confident] [tinyint] NULL,
        [Connect_Breakthru] [tinyint] NULL,
        [Connect_Relations] [tinyint] NULL,
        [Connect_Evaluation] [tinyint] NULL,
        [Connect_Annual] [tinyint] NULL,
        [Connect_Educate] [tinyint] NULL,
        [Connect_Buying] [tinyint] NULL,
        [Connect_Activities] [tinyint] NULL,
        [Connect_Resources] [tinyint] NULL,
        [Connect_Protect] [tinyint] NULL,
        [ConnectPoints] [tinyint] NULL,
        [ConnectPossiblePoints] [tinyint] NULL,
        [ConnectRatio] [float] NULL,
        [ConnectInteger] [tinyint] NULL,
        [ConnectStr] [nvarchar](255) NULL,
        [Link_Relationship] [tinyint] NULL,
        [Link_Designated] [tinyint] NULL,
        [Link_Actively] [tinyint] NULL,
        [Link_Outsourcing] [tinyint] NULL,
        [Link_Monitor] [tinyint] NULL,
        [Link_Leadership] [tinyint] NULL,
        [LinkPoint] [tinyint] NULL,
        [LinkPossiblePoints] [tinyint] NULL,
        [LinkRatio] [float] NULL,
        [LinkInteger] [tinyint] NULL,
        [LinkStr] [nvarchar](255) NULL,
        [Performance_Monitor] [tinyint] NULL,
        [Performance_Track] [tinyint] NULL,
        [Performance_Inventory] [tinyint] NULL,
        [Performance_Reward] [tinyint] NULL,
        [Performance_Financial] [tinyint] NULL,
        [Performance_Evaluate32] [tinyint] NULL,
        [Performance_Brand] [tinyint] NULL,
        [Performance_Copyright] [tinyint] NULL,
        [Performance_Climate] [tinyint] NULL,
        [Performance_Top] [tinyint] NULL,
        [Performance_Strategy] [tinyint] NULL,
        [Performance_Action] [tinyint] NULL,
        [Performance_Ratio] [tinyint] NULL,
        [Performance_Diversity] [tinyint] NULL,
        [Performance_Analysis] [tinyint] NULL,
        [Performance_Problem] [tinyint] NULL,
        [PerformancePoints] [tinyint] NULL,
        [PerformancePossiblePoints] [tinyint] NULL,
        [PerformanceRatio] [float] NULL,
        [PerformanceInteger] [tinyint] NULL,
        [PerformanceStr] [nvarchar](255) NULL,
        [OR_Oppty] [tinyint] NULL,
        [OR_External] [tinyint] NULL,
        [OR_Tolerance] [tinyint] NULL,
        [OR_Denial] [tinyint] NULL,
        [OR_Options] [tinyint] NULL,
        [OR_Divert] [tinyint] NULL,
        [OR_Innovation] [tinyint] NULL,
        [OR_Turnaround] [tinyint] NULL,
        [OR_Income10] [tinyint] NULL,
        [OR_Income5] [tinyint] NULL,
        [OR_Share10] [tinyint] NULL,
        [OR_Share5] [tinyint] NULL,
        [OR_Assets10] [tinyint] NULL,
        [OR_Assets5] [tinyint] NULL,
        [OR_LongTerm] [tinyint] NULL,
        [OR_Change] [tinyint] NULL,
        [ORPoints] [tinyint] NULL,
        [ORPossiblePoints] [tinyint] NULL,
        [ORRatio] [float] NULL,
        [ORInteger] [tinyint] NULL,
        [ORIntDiscretized] [tinyint] NULL,
        [ORStr] [nvarchar](255) NULL,
        [Ratio7Areas] [tinyint] NULL,
        [Integer7Areas] [smallint] NULL,
        	[Str7Areas] [nvarchar](255) NULL,
         CONSTRAINT [PK_tbl_DM_KM_OR_062016] PRIMARY KEY CLUSTERED
        (
        	[IP] ASC
        )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF,
        IGNORE_DUP_KEY = OFF, ALLOW_ROW_LOCKS = ON,
        ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
        ) ON [PRIMARY]

        GO
```

Fig. A8.1: Definition of the table holding questionnaire's replies.



```
/****** Script for SelectTopNRows command from SSMS  ******/
SELECT TOP 1000 [IP]
      ,[EndDate]
      ,[Sales]
      ,[Employees]
      ,[Position]
      ,[Industry]
      ,[Create_GapId]
      ,[Create_GapFix]
      ,[Create_GapSatisy]
      ,[Create_Employees]
      ,[Create_Facilities]
      ,[Create_Suggest]
      ,[Create_Experiment]
      ,[Create_Insight]
      ,[CreatePoints]
      ,[CreatePossiblePoints]
      ,[CreateRatio]
      ,[CreateInteger]
      ,[CreateStr]
      ,[Exploit_References]
      ,[Exploit_Simulate]
      ,[Exploit_Consult]
      ,[Exploit_ElectronicDB]
      ,[Exploit_Reflect]
      ,[ExploitPoints]
```

	IP	EndDate	Sales	Employees	Position	Industry	Create_GapId	Create_GapFix	Create_GapSatisy	Create_Employees	Create_Facilities	Create_Suggest	Create_Experiment	Create_Insight	CreatePoints
1	1	2014-06-25 00:00:00.000	8	7	VP/COO	Manufacturing	5	4	2	4	4	3	3	4	29
2	2	2014-07-25 00:00:00.000	8	7	CFO	Industrial Equipement Manufacturing	5	4	2	5	4	5	5	3	33
3	3	2014-03-19 00:00:00.000	8	7	President	Construction	3	2	2	2	4	2	3	5	23
4	4	2014-04-29 00:00:00.000	4	6	CEO	Auto Glass Repair	3	3	3	5	4	2	1	5	26
5	5	2014-03-25 00:00:00.000	8	7	Manager IT and Process Automation	Metals, Mining, and Manufacturing	4	4	3	4	5	3	3	5	31
6	6	2014-03-21 00:00:00.000	8	7	President/CEO	Electric Utility	4	2	2	4	4	4	4	5	29
7	7	2014-03-20 00:00:00.000	8	7	CEO	Contact center and fulfillment solutions	4	4	2	4	5	3	4	4	30
8	8	2014-03-19 00:00:00.000	8	7	CEO	Healthcare/Hospital	4	4	2	5	3	3	4	5	30
9	9	2014-03-11 00:00:00.000	8	7	VP	Healthcare	4	4	3	3	4	2	2	4	26
10	10	2014-03-03 00:00:00.000	8	7	CEO	Agriculture	4	2	2	5	5	2	4	5	29
11	11	2014-04-28 00:00:00.000	4	4	CFO	Recreation equipment manufacturer.	4	4	4	4	5	2	4	4	31
12	12	2014-04-28 00:00:00.000	7	5	Controller	Oil Refinery	4	4	2	4	4	2	4	4	28
13	13	2014-04-28 00:00:00.000	5	6	CEO	Medical Supply	4	4	2	3	4	2	5	4	28
14	14	2014-06-09 16:38:22.000	8	7	CFO	Cabinet Manufacturer	4	4	2	4	4	4	4	4	30
15	15	2014-06-06 16:50:46.000	8	7	President/CEO	Agriculture/Energy	4	4	2	3	4	2	2	5	26
16	16	2014-06-05 02:29:30.000	8	7	Coo	Iowa	4	5	2	4	4	2	3	4	28
17	17	2014-06-04 21:57:52.000	8	7	Medical Director	Medical Management/Health Insurance	4	3	1	2	4	3	2	4	23
18	18	2014-06-04 20:52:05.000	8	7	CEO	Telecommunications	4	4	3	5	4	4	3	5	32

Query executed successfully.

ANALYTICS47\ANALYTICS01 (11... | ANALYTICS47\Michael (56) | master | 00:00:00 | 45 rows

Fig. A8.2: Presentation of the relational table holding the questionnaire's data along in the Microsoft SQL Server Management Studio

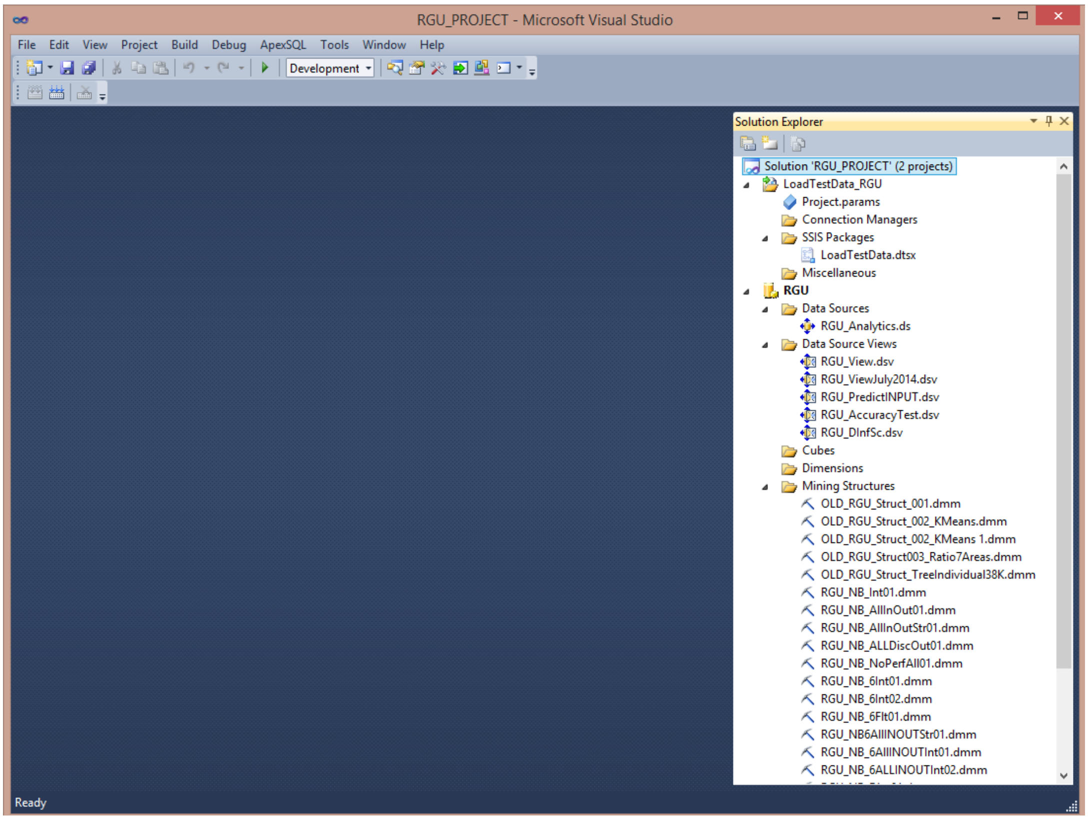
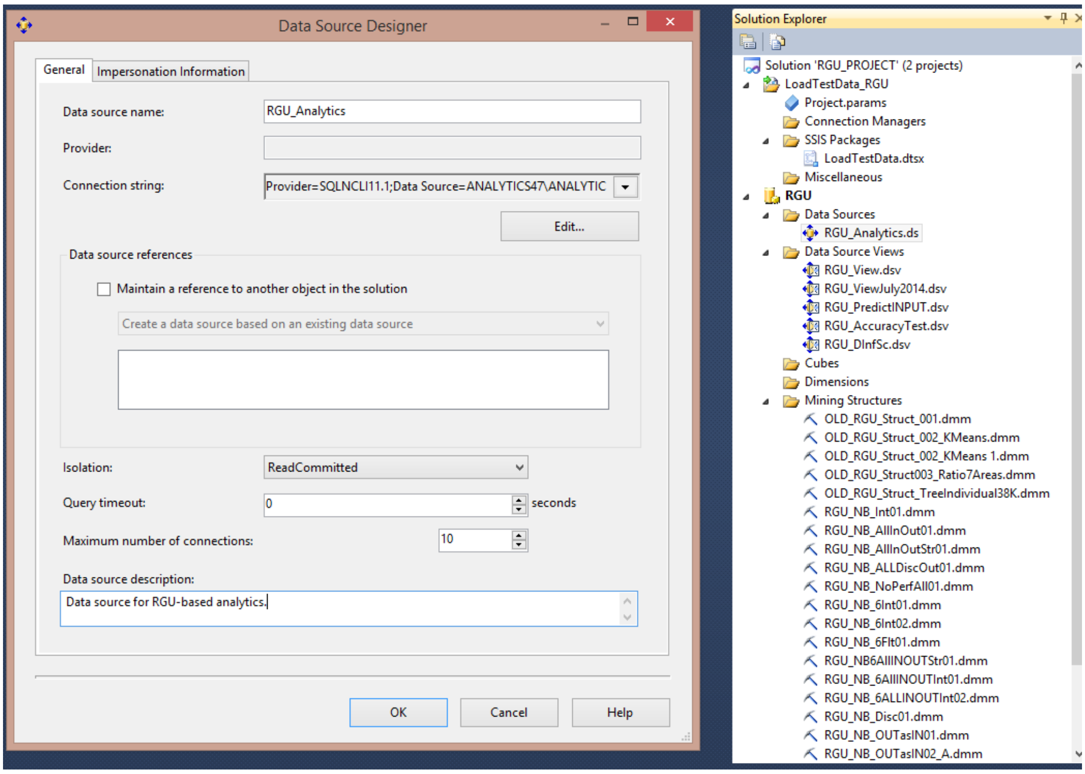


Fig. A8.3: Visual Studio development environment



Data source name: RGU_Analytics

Provider:

Connection string: Provider=SQLNCLI11.1;Data Source=ANALYTICS47\ANALYTIC

Data source references

□ Maintain a reference to another object in the solution

Create a data source based on an existing data source

Isolation: ReadCommitted

Query timeout: 0 seconds

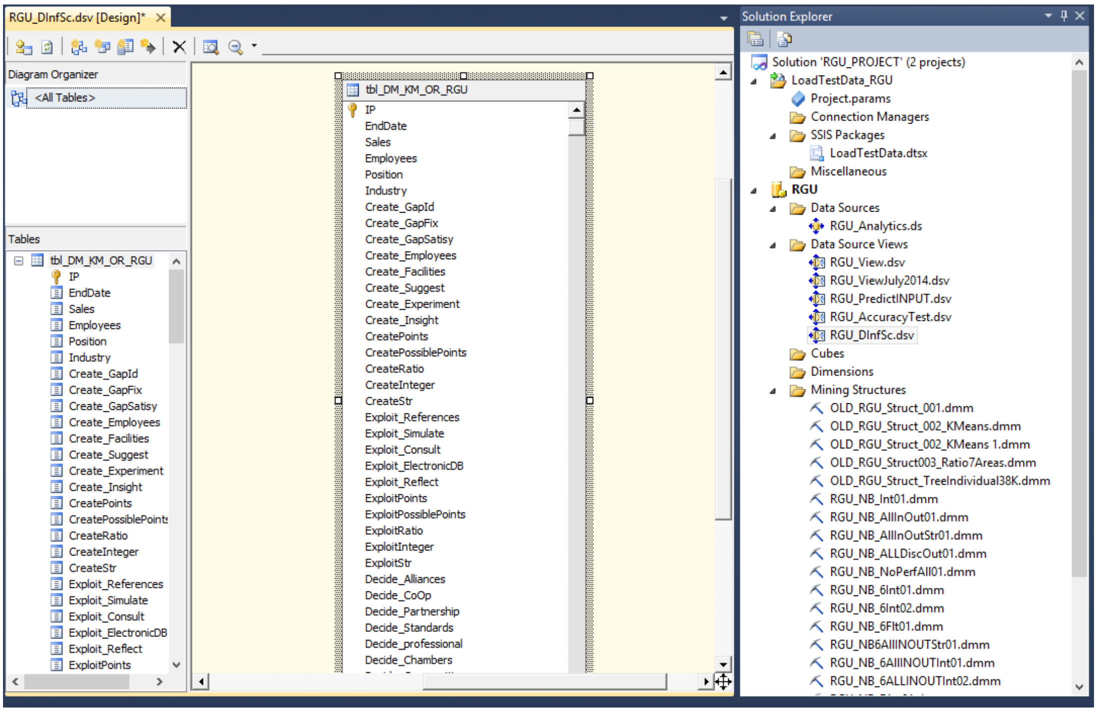
Maximum number of connections: 10

Data source description:

Data source for RGU-based analytics.

Solution 'RGU_PROJECT' (2 projects)
- LoadTestData_RGU
  - Project.params
  - Connection Managers
  - SSIS Packages
    - LoadTestData.dtsx
  - Miscellaneous
- RGU
  - Data Sources
    - RGU_Analytics.ds
  - Data Source Views
    - RGU_View.dsv
    - RGU_ViewJuly2014.dsv
    - RGU_PredictINPUT.dsv
    - RGU_AccuracyTest.dsv
    - RGU_DInfSc.dsv
  - Cubes
  - Dimensions
  - Mining Structures
    - OLD_RGU_Struct_001.dmm
    - OLD_RGU_Struct_002_KMeans.dmm
    - OLD_RGU_Struct_002_KMeans 1.dmm
    - OLD_RGU_Struct003_Ratio7Areas.dmm
    - OLD_RGU_Struct_TreeIndividual38K.dmm
    - RGU_NB_Int01.dmm
    - RGU_NB_AllInOut01.dmm
    - RGU_NB_AllInOutStr01.dmm
    - RGU_NB_ALLDiscOut01.dmm
    - RGU_NB_NoPerfAll01.dmm
    - RGU_NB_6Int01.dmm
    - RGU_NB_6Int02.dmm
    - RGU_NB_6Flt01.dmm
    - RGU_NB6AllINOUTStr01.dmm
    - RGU_NB_6AllINOUTInt01.dmm
    - RGU_NB_6ALLINOUTInt02.dmm
    - RGU_NB_Disc01.dmm
    - RGU_NB_OUTasIN01.dmm
    - RGU_NB_OUTasIN02_A.dmm

Fig. A8.4: Data source (database level)

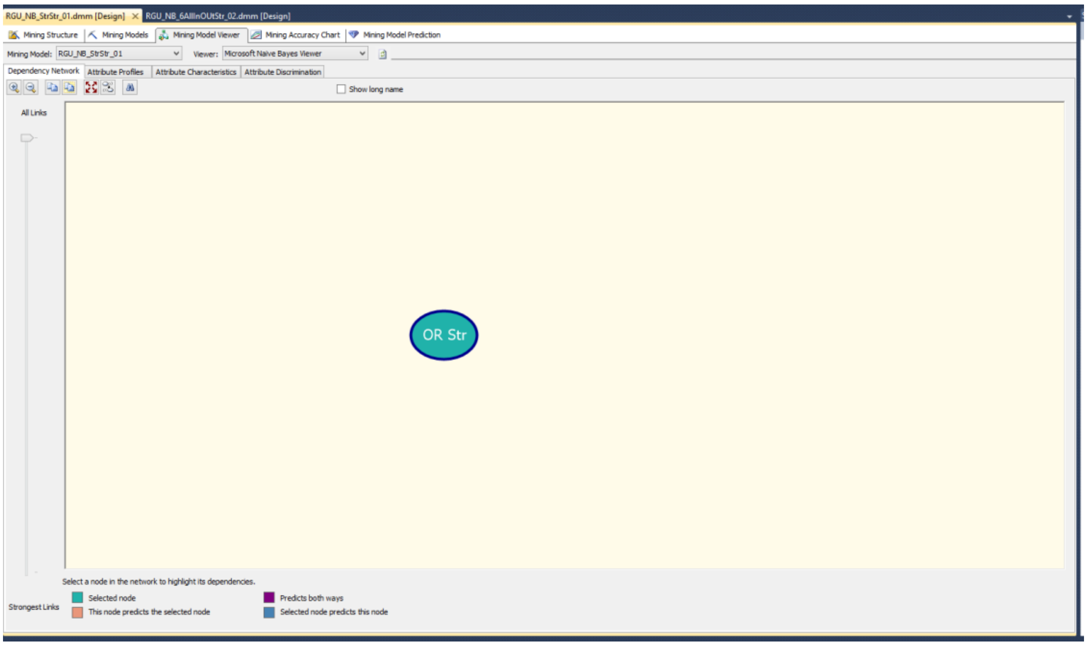


tbl_DM_KM_OR_RGU
- IP
- EndDate
- Sales
- Employees
- Position
- Industry
- Create_GapId
- Create_GapFix
- Create_GapSatisy
- Create_Employees
- Create_Facilities
- Create_Suggest
- Create_Experiment
- Create_Insight
- CreatePoints
- CreatePossiblePoints
- CreateRatio
- CreateInteger
- CreateStr
- Exploit_References
- Exploit_Simulate
- Exploit_Consult
- Exploit_ElectronicDB
- Exploit_Reflect
- ExploitPoints
- ExploitPossiblePoints
- ExploitRatio
- ExploitInteger
- ExploitStr
- Decide_Alliances
- Decide_CoOp
- Decide_Partnership
- Decide_Standards
- Decide_professional
- Decide_Chambers

tbl_DM_KM_OR_RGU
IP
EndDate
Sales
Employees
Position
Industry
Create_GapId
Create_GapFix
Create_GapSatisy
Create_Employees
Create_Facilities
Create_Suggest
Create_Experiment
Create_Insight
CreatePoints
CreatePossiblePoints
CreateRatio
CreateInteger
CreateStr
Exploit_References
Exploit_Simulate
Exploit_Consult
Exploit_ElectronicDB
Exploit_Reflect
ExploitPoints

Solution 'RGU_PROJECT' (2 projects)
LoadTestData_RGU
Project.params
Connection Managers
SSIS Packages
LoadTestData.dtsx
Miscellaneous
RGU
Data Sources
RGU_Analytics.ds
Data Source Views
RGU_View.dsv
RGU_ViewJuly2014.dsv
RGU_PredictINPUT.dsv
RGU_AccuracyTest.dsv
RGU_DInfSc.dsv
Cubes
Dimensions
Mining Structures
OLD_RGU_Struct_001.dmm
OLD_RGU_Struct_002_KMeans.dmm
OLD_RGU_Struct_002_KMeans 1.dmm
OLD_RGU_Struct003_Ratio7Areas.dmm
OLD_RGU_Struct_TreeIndividual38K.dmm
RGU_NB_Int01.dmm
RGU_NB_AllInOut01.dmm
RGU_NB_AllInOutStr01.dmm
RGU_NB_ALLDiscOut01.dmm
RGU_NB_NoPerfAll01.dmm
RGU_NB_6Int01.dmm
RGU_NB_6Int02.dmm
RGU_NB_6Flt01.dmm
RGU_NB6AllINOUTStr01.dmm
RGU_NB_6AllINOUTInt01.dmm
RGU_NB_6ALLINOUTInt02.dmm

Fig. A8.5: Data source view (table level)



Mining Model: RGU_NB_StrStr_01

Viewer: Microsoft Naive Bayes Viewer

□ Show long name

Select a node in the network to highlight its dependencies.

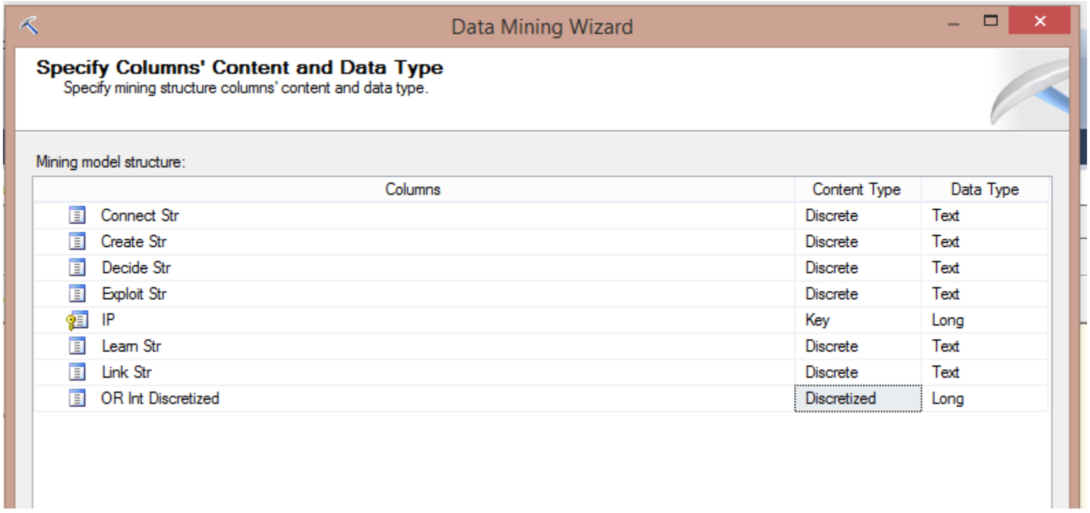
🟩 Selected node

🟥 This node predicts the selected node

🟪 Predicts both ways

🟦 Selected node predicts this node

Fig. A8.6: Outcome (Dependency Network) of DM Naïve Bayes model using input and output of categorical (string) type



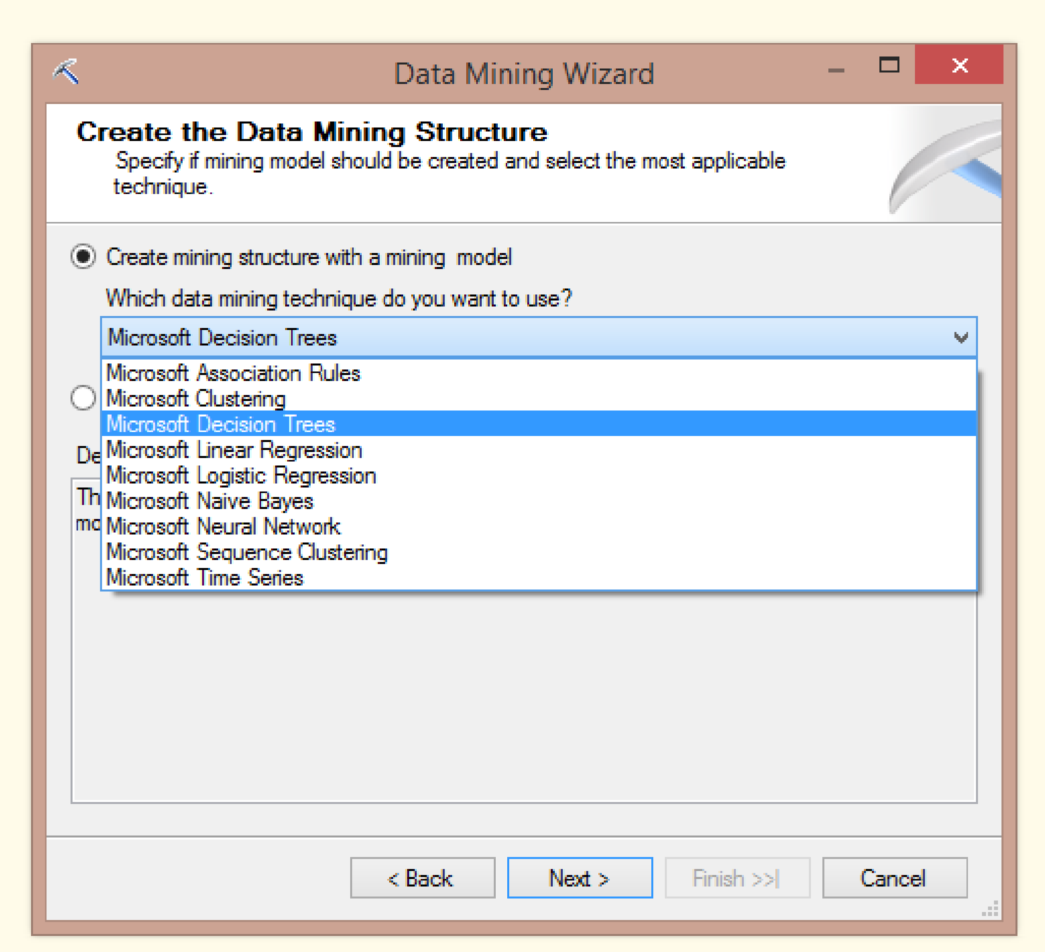
Specify Columns' Content and Data Type

Specify mining structure columns' content and data type.

Mining model structure:

Columns	Content Type	Data Type
Connect Str	Discrete	Text
Create Str	Discrete	Text
Decide Str	Discrete	Text
Exploit Str	Discrete	Text
IP	Key	Long
Learn Str	Discrete	Text
Link Str	Discrete	Text
OR Int Discretized	Discretized	Long

Fig. A8.7: Setting the content type of a DM model



Create the Data Mining Structure

Specify if mining model should be created and select the most applicable technique.

● Create mining structure with a mining model

Which data mining technique do you want to use?

Microsoft Decision Trees

- Microsoft Association Rules
- Microsoft Clustering
- Microsoft Decision Trees
- Microsoft Linear Regression
- Microsoft Logistic Regression
- Microsoft Naive Bayes
- Microsoft Neural Network
- Microsoft Sequence Clustering
- Microsoft Time Series

○

De

Th
mo

Fig. A8.8: Data mining techniques available in MS SQL Server 2012

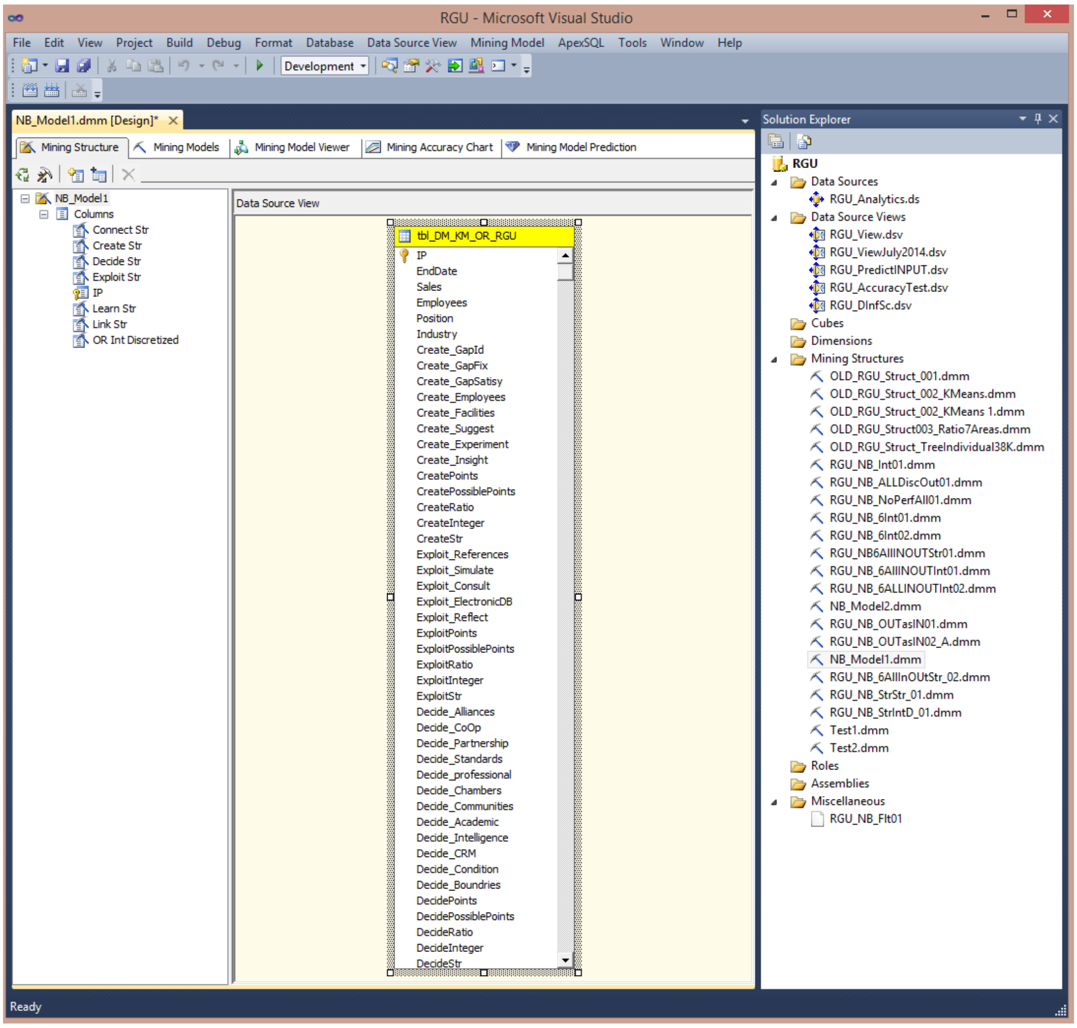
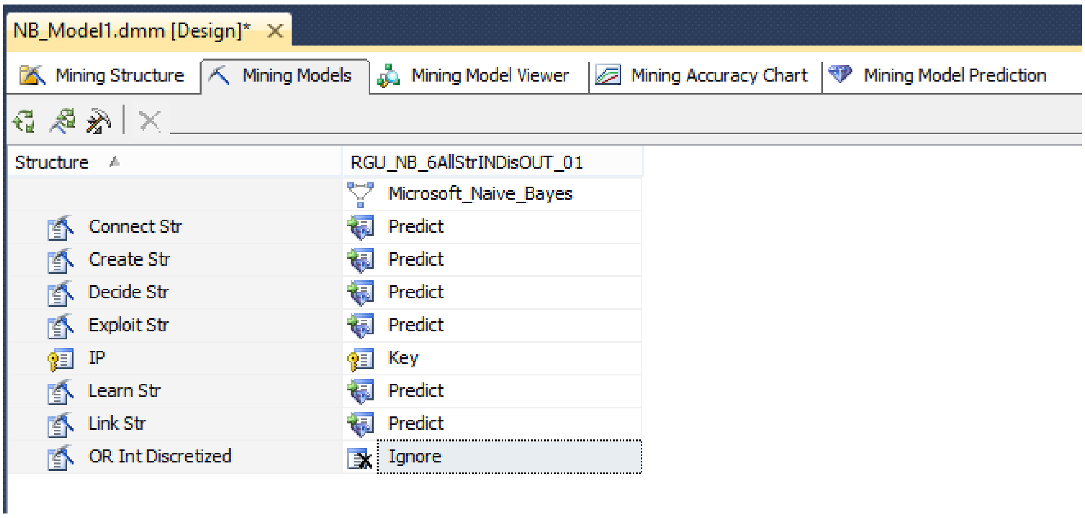
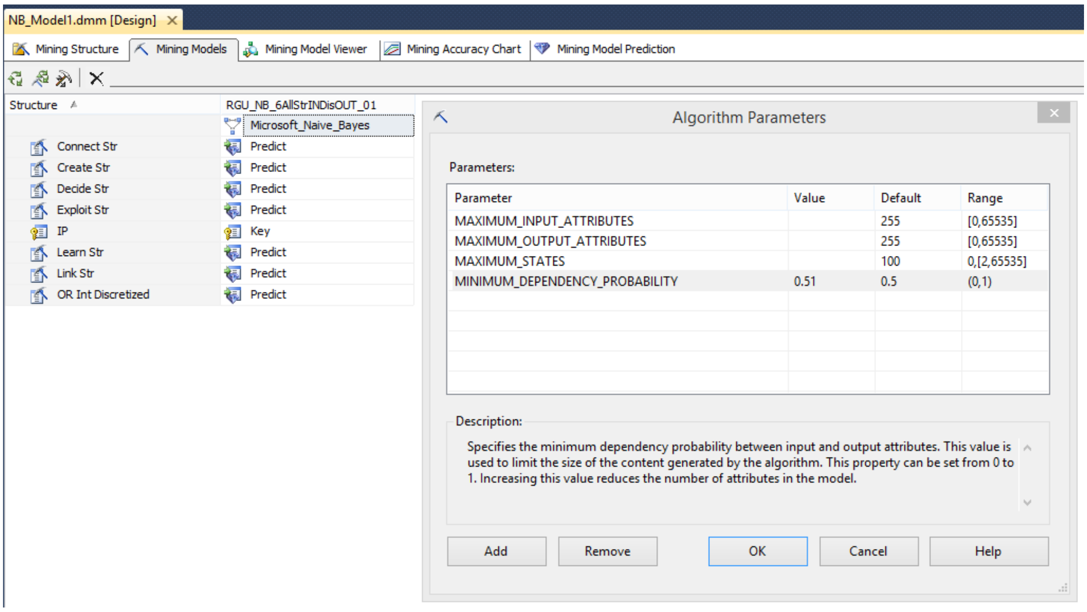


Fig. A8.9: NB_Model1 structure



Structure	RGU_NB_6AllStrINDisOUT_01
	Microsoft_Naive_Bayes
Connect Str	Predict
Create Str	Predict
Decide Str	Predict
Exploit Str	Predict
IP	Key
Learn Str	Predict
Link Str	Predict
OR Int Discretized	Ignore

Fig. A8.10: NB_Model1 – mining model



Structure	RGU_NB_6AllStrINDisOUT_01
	Microsoft_Naive_Bayes
Connect Str	Predict
Create Str	Predict
Decide Str	Predict
Exploit Str	Predict
IP	Key
Learn Str	Predict
Link Str	Predict
OR Int Discretized	Predict

Algorithm Parameters

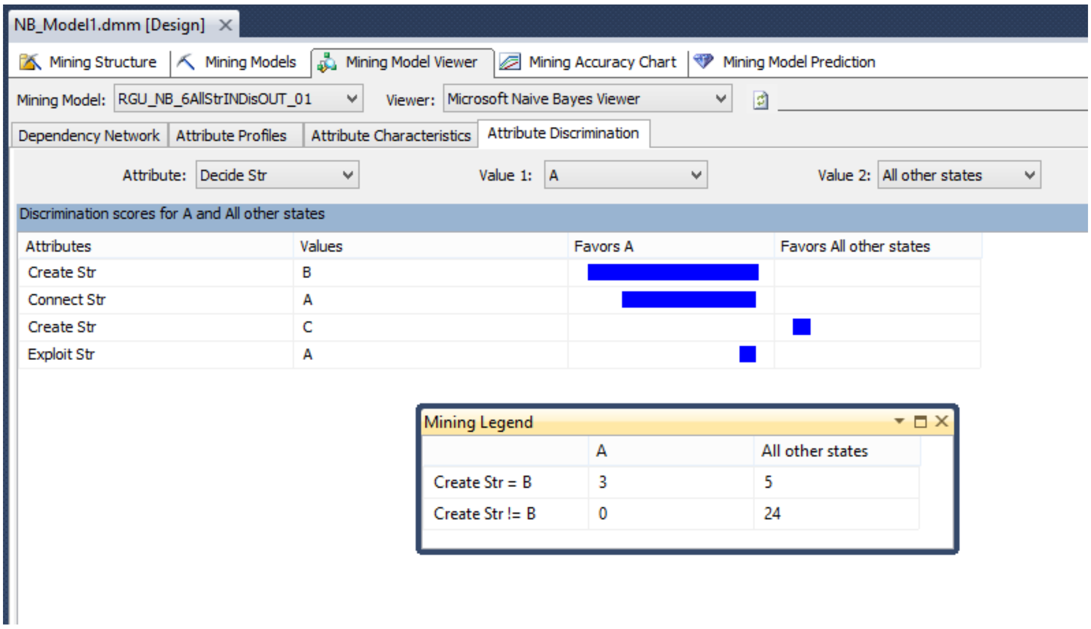
Parameters:

Parameter	Value	Default	Range
MAXIMUM_INPUT_ATTRIBUTES		255	[0,65535]
MAXIMUM_OUTPUT_ATTRIBUTES		255	[0,65535]
MAXIMUM_STATES		100	0,[2,65535]
MINIMUM_DEPENDENCY_PROBABILITY	0.51	0.5	(0,1)

Description:

Specifies the minimum dependency probability between input and output attributes. This value is used to limit the size of the content generated by the algorithm. This property can be set from 0 to 1. Increasing this value reduces the number of attributes in the model.

Fig. A8.11: NB algorithm's parameters



Mining Model: RGU_NB_6AllStrINDisOUT_01 Viewer: Microsoft Naive Bayes Viewer

Attribute: Decide Str Value 1: A Value 2: All other states

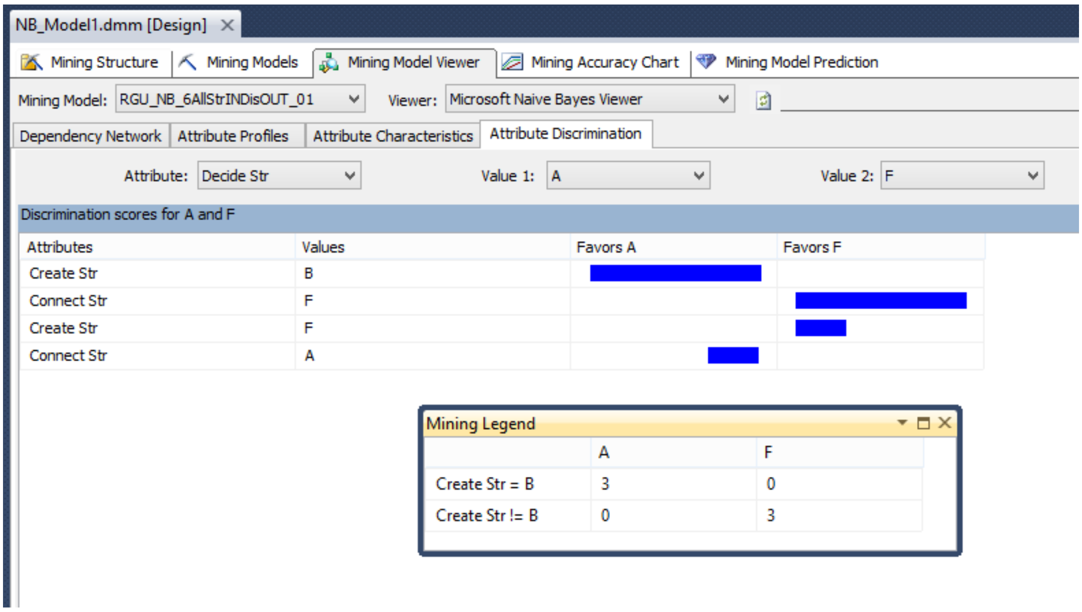
Discrimination scores for A and All other states

Attributes	Values	Favors A	Favors All other states
Create Str	B	■	
Connect Str	A	■	
Create Str	C		■
Exploit Str	A	■	

Mining Legend

	A	All other states
Create Str = B	3	5
Create Str != B	0	24

Fig. A8.12: Example of attribute discrimination: 'Decide Str' value of 'A' vs. 'All other states'



Mining Model: RGU_NB_6AllStrINDisOUT_01 Viewer: Microsoft Naive Bayes Viewer

Attribute: Decide Str Value 1: A Value 2: F

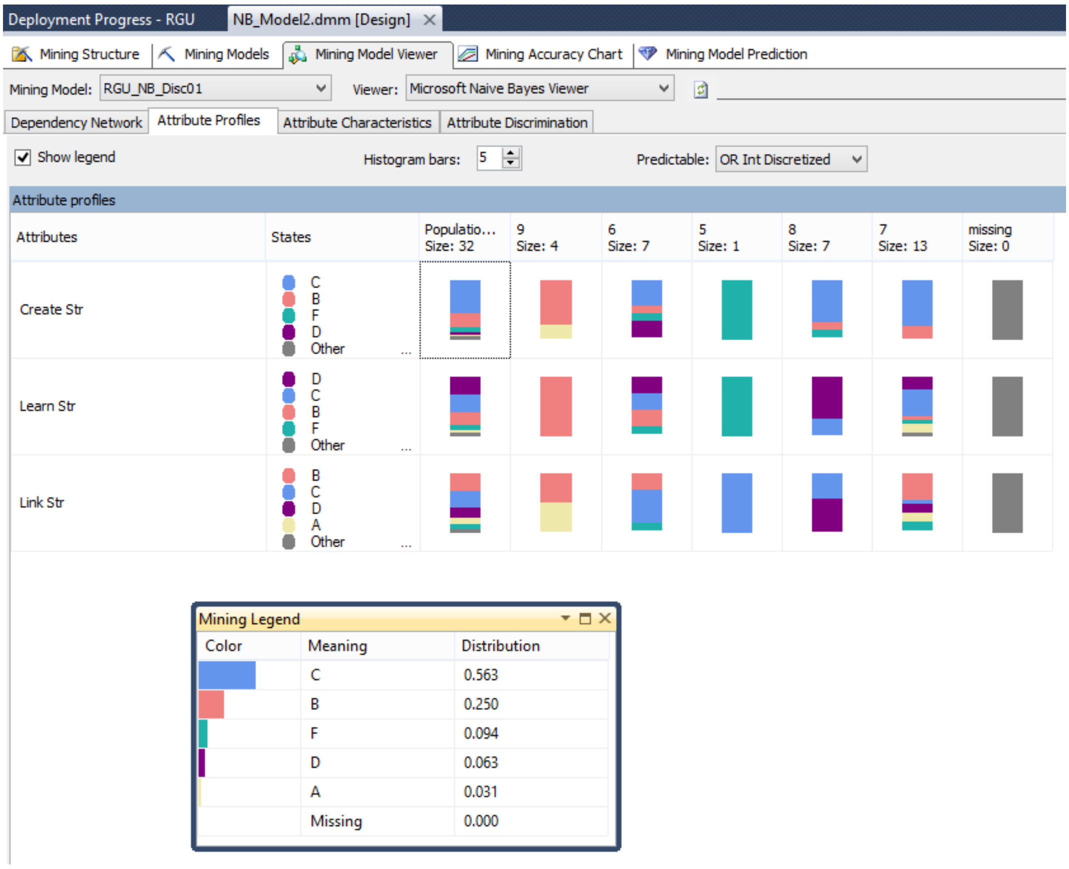
Discrimination scores for A and F

Attributes	Values	Favors A	Favors F
Create Str	B	■	
Connect Str	F		■
Create Str	F		■
Connect Str	A	■	

Mining Legend

	A	F
Create Str = B	3	0
Create Str != B	0	3

Fig. A8.13: Example of attribute discrimination: 'Decide Str' value of 'A' vs. 'F'



Mining Model: RGU_NB_Disc01 Viewer: Microsoft Naive Bayes Viewer

☑ Show legend Histogram bars: 5 Predictable: OR Int Discretized

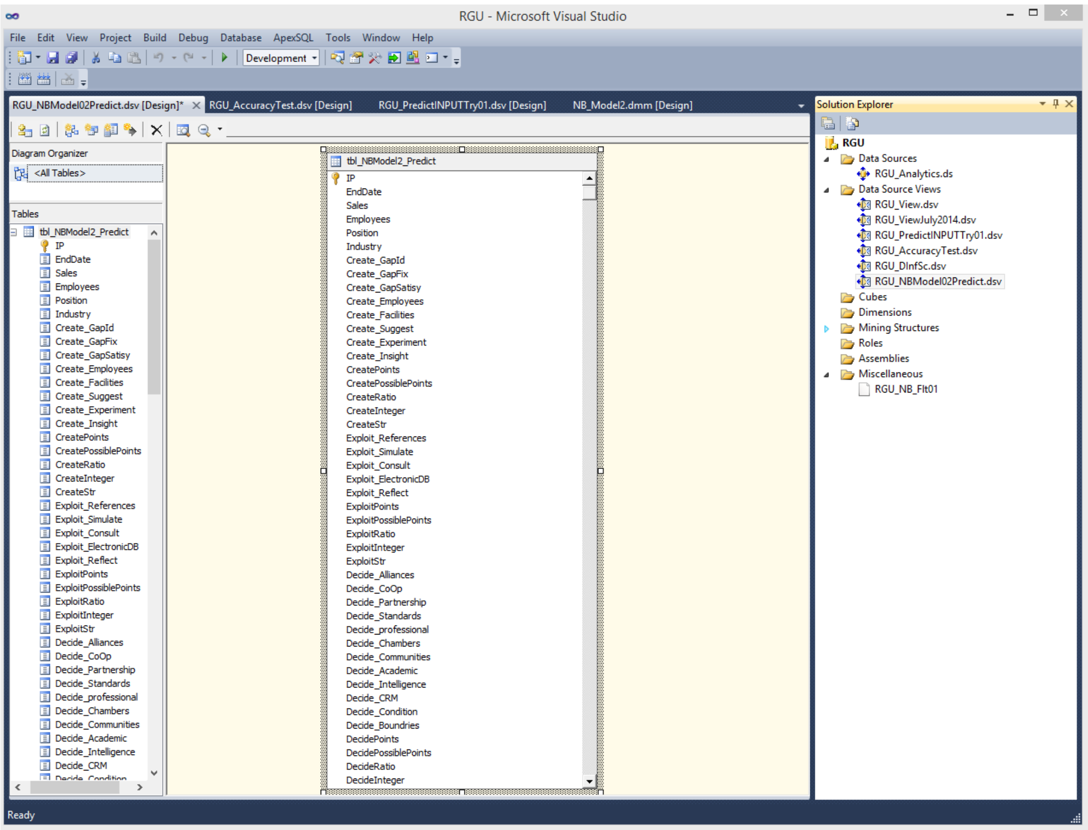
Attribute profiles

Attributes	States	Populatio... Size: 32	9 Size: 4	6 Size: 7	5 Size: 1	8 Size: 7	7 Size: 13	missing Size: 0
Create Str	C B F D Other ...							
Learn Str	D C B F Other ...							
Link Str	B C D A Other ...							

Mining Legend

Color	Meaning	Distribution
🟦	C	0.563
🟥	B	0.250
🟩	F	0.094
🟪	D	0.063
🟨	A	0.031
	Missing	0.000

Fig. A8.14: Example of Mining Legend providing statistical information with regards to the (selected) population composition associated with the 'Create Str' input variable

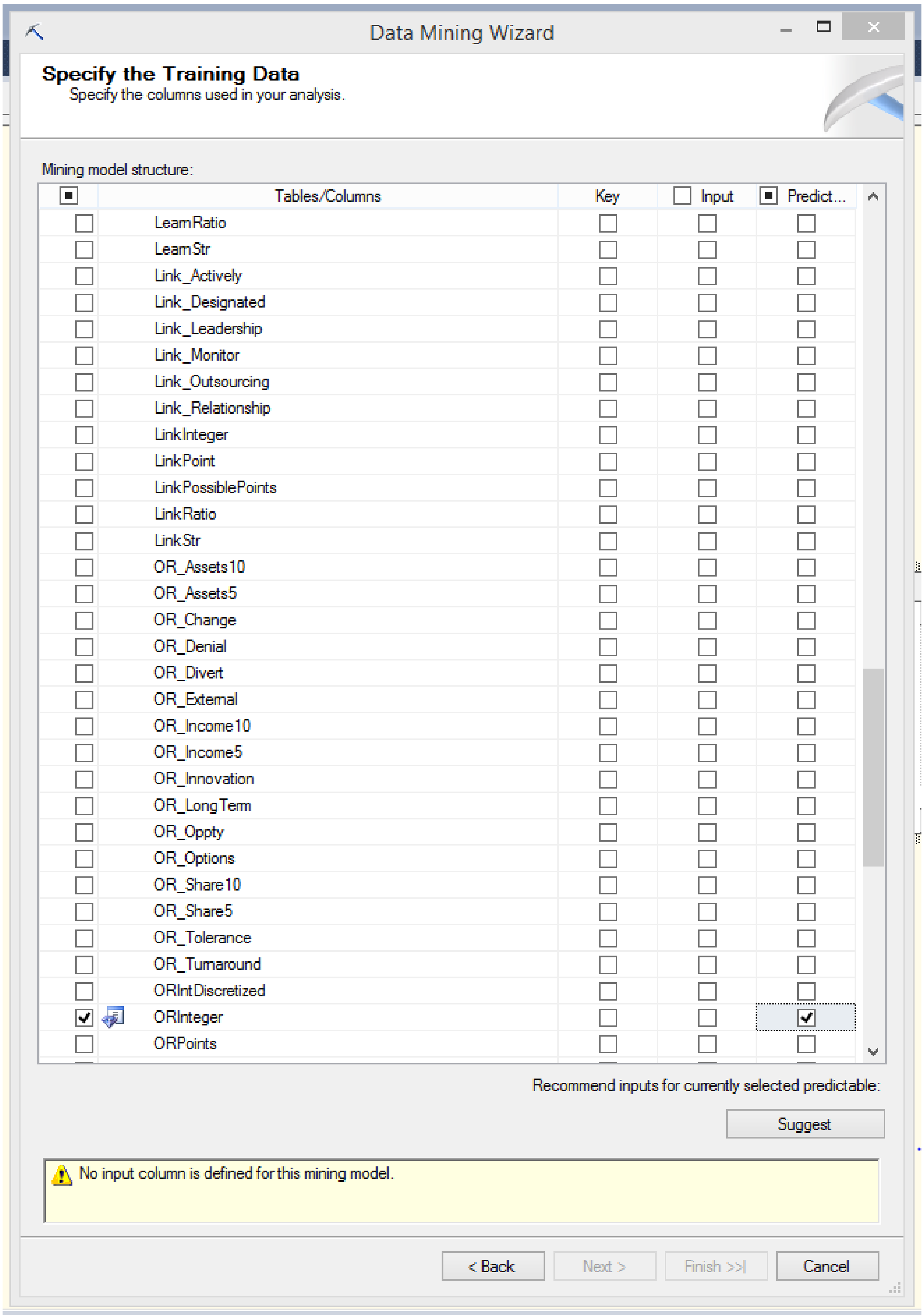

RGU - Microsoft Visual Studio
File Edit View Project Build Debug Database ApexSQL Tools Window Help
Development
RGU_NBModel02Predict.dsv [Design]*
RGU_AccuracyTest.dsv [Design]
RGU_PredictINPUTTry01.dsv [Design]
NB_Model2.dmm [Design]
Diagram Organizer
<All Tables>
Tables
Solution Explorer
Ready

tbl_NBModel2_Predict

IP
EndDate
Sales
Employees
Position
Industry
Create_GapId
Create_GapFix
Create_GapSatisy
Create_Employees
Create_Facilties
Create_Suggest
Create_Experiment
Create_Insight
CreatePoints
CreatePossiblePoints
CreateRatio
CreateInteger
CreateStr
Exploit_References
Exploit_Simulate
Exploit_Consult
Exploit_ElectronicDB
Exploit_Reflect
ExploitPoints
ExploitPossiblePoints
ExploitRatio
ExploitInteger
ExploitStr
Decide_Alliances
Decide_CoOp
Decide_Partnership
Decide_Standards
Decide_professional
Decide_Chambers
Decide_Communities
Decide_Academic
Decide_Intelligence
Decide_CRM
Decide_Condition
Decide_Boundries
DecidePoints
DecidePossiblePoints
DecideRatio
DecideInteger

RGU
- Data Sources
  - RGU_Analytics.ds
- Data Source Views
  - RGU_View.dsv
  - RGU_ViewJuly2014.dsv
  - RGU_PredictINPUTTry01.dsv
  - RGU_AccuracyTest.dsv
  - RGU_DInfSc.dsv
  - RGU_NBModel02Predict.dsv
- Cubes
- Dimensions
- Mining Structures
- Roles
- Assemblies
- Miscellaneous
  - RGU_NB_FIt01

Fig. A8.15: Illustration of the prediction construct for NB_Model02



Specify the Training Data

Specify the columns used in your analysis.

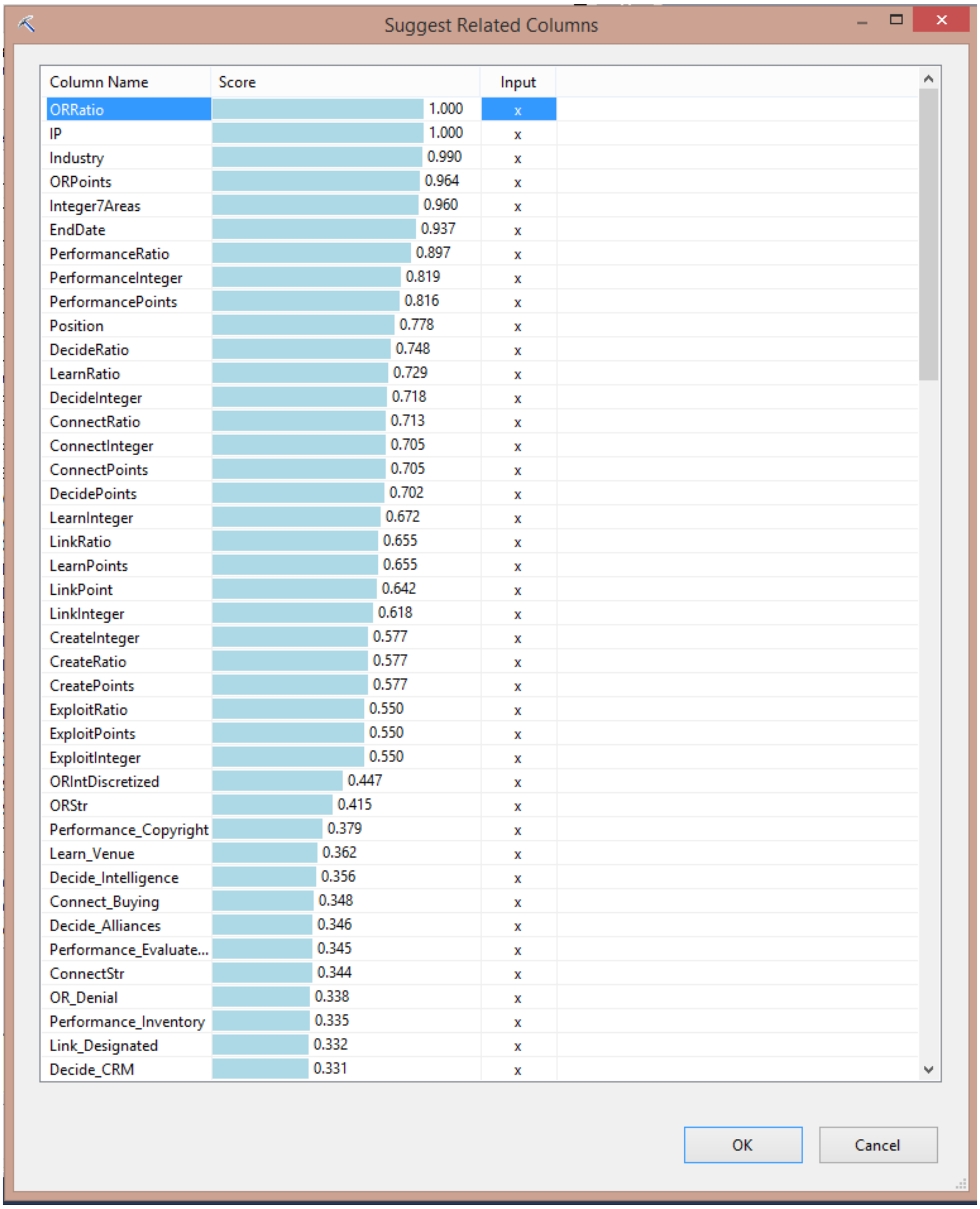
Mining model structure:

■	Tables/Columns	Key	□ Input	■ Predict...
□	LearnRatio	□	□	□
□	LearnStr	□	□	□
□	Link_Actively	□	□	□
□	Link_Designated	□	□	□
□	Link_Leadership	□	□	□
□	Link_Monitor	□	□	□
□	Link_Outsourcing	□	□	□
□	Link_Relationship	□	□	□
□	LinkInteger	□	□	□
□	LinkPoint	□	□	□
□	LinkPossiblePoints	□	□	□
□	LinkRatio	□	□	□
□	LinkStr	□	□	□
□	OR_Assets10	□	□	□
□	OR_Assets5	□	□	□
□	OR_Change	□	□	□
□	OR_Denial	□	□	□
□	OR_Divert	□	□	□
□	OR_External	□	□	□
□	OR_Income10	□	□	□
□	OR_Income5	□	□	□
□	OR_Innovation	□	□	□
□	OR_LongTerm	□	□	□
□	OR_Oppty	□	□	□
□	OR_Options	□	□	□
□	OR_Share10	□	□	□
□	OR_Share5	□	□	□
□	OR_Tolerance	□	□	□
□	OR_Turnaround	□	□	□
□	ORIntDiscretized	□	□	□
☑	ORInteger	□	□	☑
□	ORPoints	□	□	□

Recommend inputs for currently selected predictable:

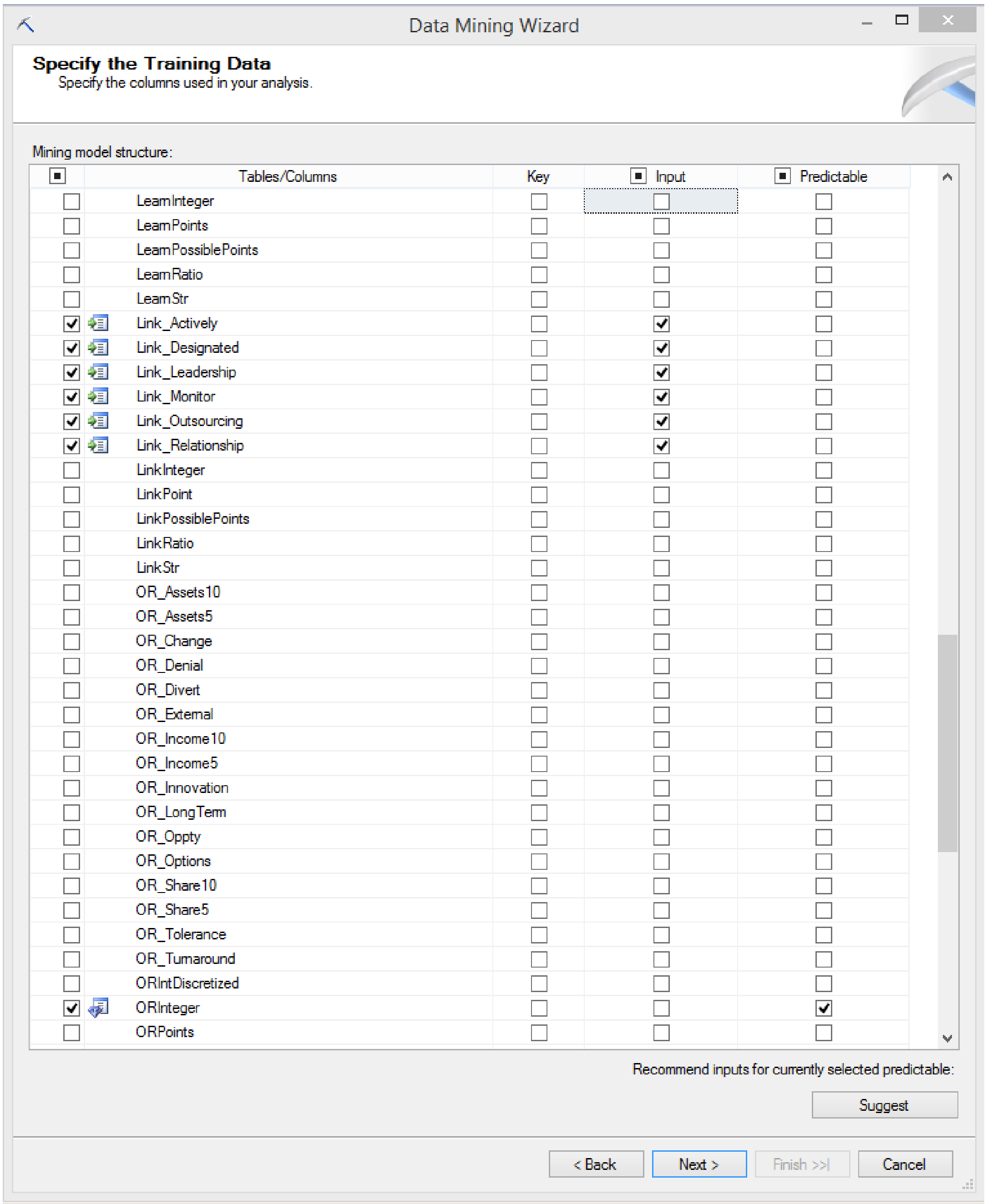
⚠ No input column is defined for this mining model.

Fig. A8.16: Illustration of availability of the 'Suggest' button (to suggest input variables) upon selection of the output variable



Column Name	Score	Input
ORRatio	1.000	x
IP	1.000	x
Industry	0.990	x
ORPoints	0.964	x
Integer7Areas	0.960	x
EndDate	0.937	x
PerformanceRatio	0.897	x
PerformanceInteger	0.819	x
PerformancePoints	0.816	x
Position	0.778	x
DecideRatio	0.748	x
LearnRatio	0.729	x
DecideInteger	0.718	x
ConnectRatio	0.713	x
ConnectInteger	0.705	x
ConnectPoints	0.705	x
DecidePoints	0.702	x
LearnInteger	0.672	x
LinkRatio	0.655	x
LearnPoints	0.655	x
LinkPoint	0.642	x
LinkInteger	0.618	x
CreateInteger	0.577	x
CreateRatio	0.577	x
CreatePoints	0.577	x
ExploitRatio	0.550	x
ExploitPoints	0.550	x
ExploitInteger	0.550	x
ORIntDiscretized	0.447	x
ORStr	0.415	x
Performance_Copyright	0.379	x
Learn_Venue	0.362	x
Decide_Intelligence	0.356	x
Connect_Buying	0.348	x
Decide_Alliances	0.346	x
Performance_Evaluate...	0.345	x
ConnectStr	0.344	x
OR_Denial	0.338	x
Performance_Inventory	0.335	x
Link_Designated	0.332	x
Decide_CRM	0.331	x

Fig. A8.17: Some suggested by the DM wizard choices of input variables



Specify the Training Data

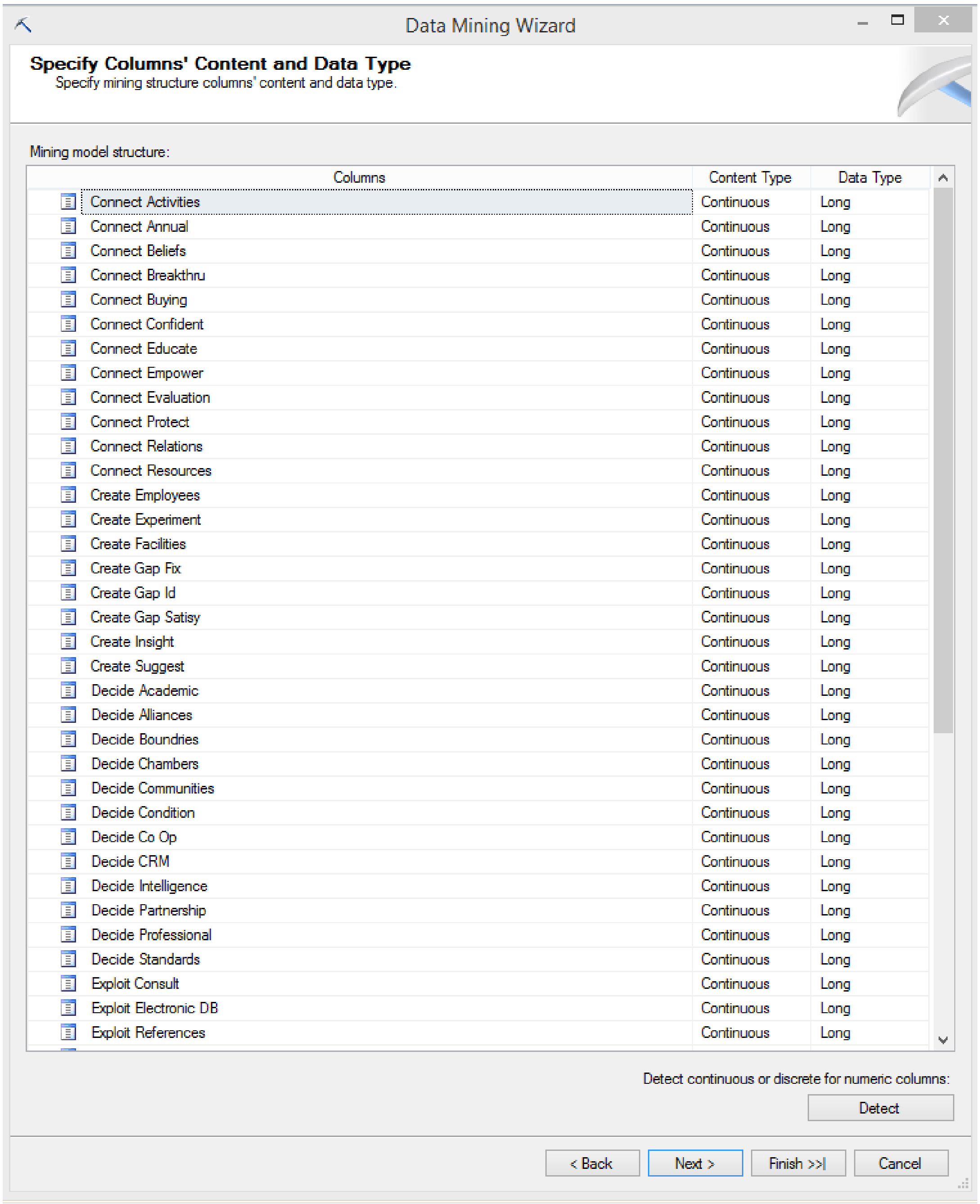
Specify the columns used in your analysis.

Mining model structure:

■	Tables/Columns	Key	■ Input	■ Predictable
□	LearnInteger	□	□	□
□	LearnPoints	□	□	□
□	LearnPossiblePoints	□	□	□
□	LearnRatio	□	□	□
□	LearnStr	□	□	□
☑	Link_Actively	□	☑	□
☑	Link_Designated	□	☑	□
☑	Link_Leadership	□	☑	□
☑	Link_Monitor	□	☑	□
☑	Link_Outsourcing	□	☑	□
☑	Link_Relationship	□	☑	□
□	LinkInteger	□	□	□
□	LinkPoint	□	□	□
□	LinkPossiblePoints	□	□	□
□	LinkRatio	□	□	□
□	LinkStr	□	□	□
□	OR_Assets10	□	□	□
□	OR_Assets5	□	□	□
□	OR_Change	□	□	□
□	OR_Denial	□	□	□
□	OR_Divert	□	□	□
□	OR_External	□	□	□
□	OR_Income10	□	□	□
□	OR_Income5	□	□	□
□	OR_Innovation	□	□	□
□	OR_LongTerm	□	□	□
□	OR_Oppty	□	□	□
□	OR_Options	□	□	□
□	OR_Share10	□	□	□
□	OR_Share5	□	□	□
□	OR_Tolerance	□	□	□
□	OR_Turnaround	□	□	□
□	ORIntDiscretized	□	□	□
☑	ORInteger	□	□	☑
□	ORPoints	□	□	□

Recommend inputs for currently selected predictable:

Fig. A8.18: Illustration of selection of output and some input variables



Specify Columns' Content and Data Type

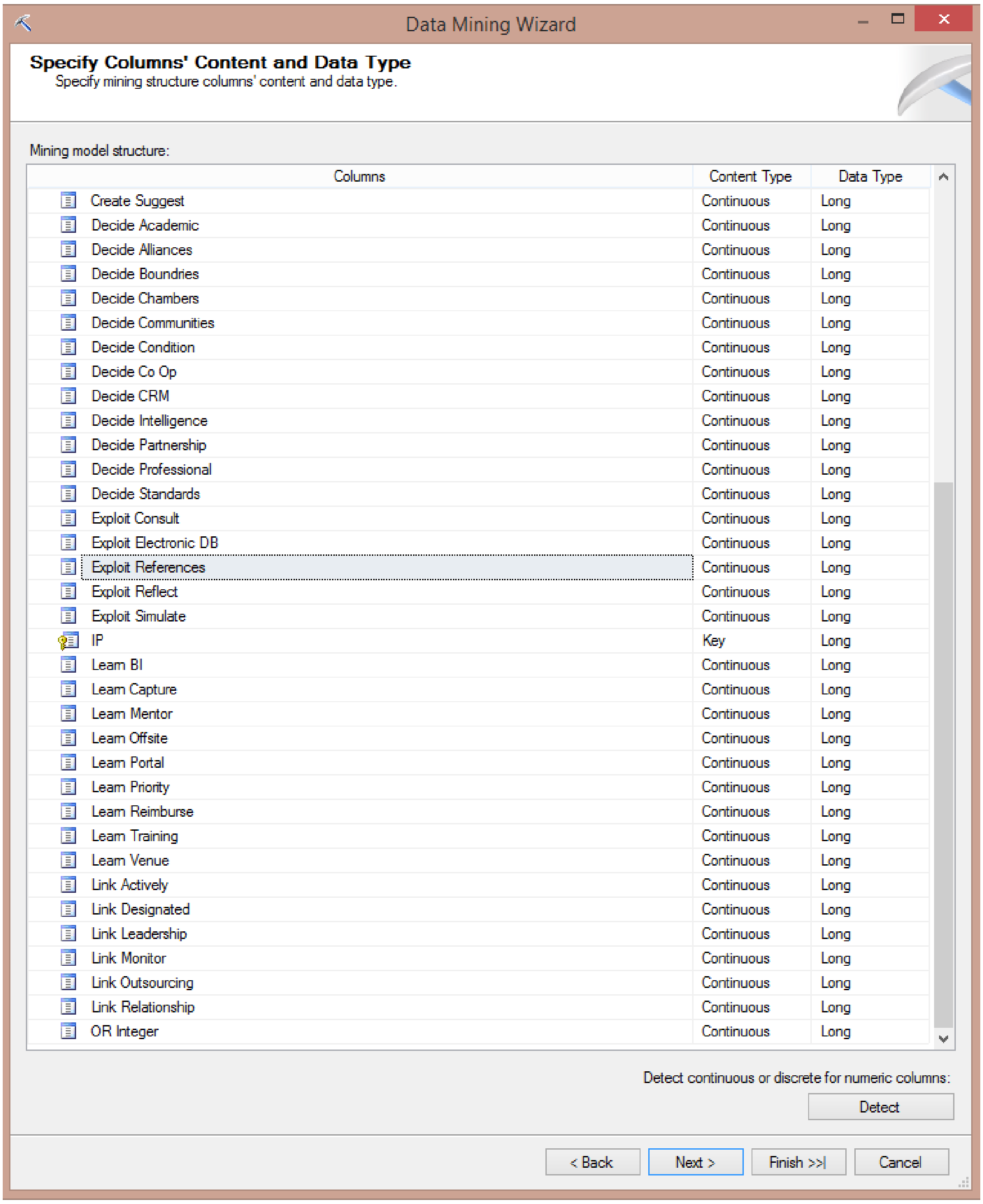
Specify mining structure columns' content and data type.

Mining model structure:

Columns	Content Type	Data Type
Connect Activities	Continuous	Long
Connect Annual	Continuous	Long
Connect Beliefs	Continuous	Long
Connect Breakthru	Continuous	Long
Connect Buying	Continuous	Long
Connect Confident	Continuous	Long
Connect Educate	Continuous	Long
Connect Empower	Continuous	Long
Connect Evaluation	Continuous	Long
Connect Protect	Continuous	Long
Connect Relations	Continuous	Long
Connect Resources	Continuous	Long
Create Employees	Continuous	Long
Create Experiment	Continuous	Long
Create Facilities	Continuous	Long
Create Gap Fix	Continuous	Long
Create Gap Id	Continuous	Long
Create Gap Satisy	Continuous	Long
Create Insight	Continuous	Long
Create Suggest	Continuous	Long
Decide Academic	Continuous	Long
Decide Alliances	Continuous	Long
Decide Boundries	Continuous	Long
Decide Chambers	Continuous	Long
Decide Communities	Continuous	Long
Decide Condition	Continuous	Long
Decide Co Op	Continuous	Long
Decide CRM	Continuous	Long
Decide Intelligence	Continuous	Long
Decide Partnership	Continuous	Long
Decide Professional	Continuous	Long
Decide Standards	Continuous	Long
Exploit Consult	Continuous	Long
Exploit Electronic DB	Continuous	Long
Exploit References	Continuous	Long

Detect continuous or discrete for numeric columns:

Fig. A8.19: List of input variables used in model Cluster_Model1 – part 1



Specify Columns' Content and Data Type

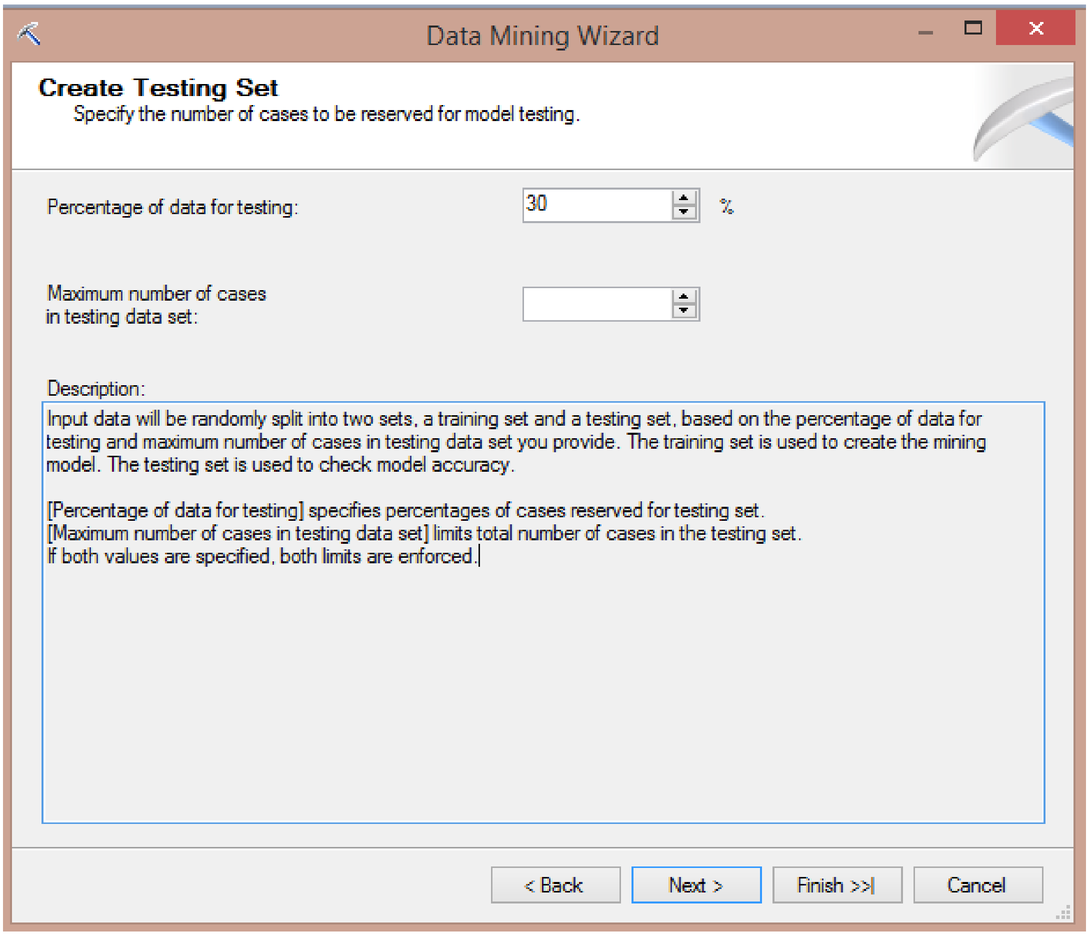
Specify mining structure columns' content and data type.

Mining model structure:

Columns	Content Type	Data Type
Create Suggest	Continuous	Long
Decide Academic	Continuous	Long
Decide Alliances	Continuous	Long
Decide Boundries	Continuous	Long
Decide Chambers	Continuous	Long
Decide Communities	Continuous	Long
Decide Condition	Continuous	Long
Decide Co Op	Continuous	Long
Decide CRM	Continuous	Long
Decide Intelligence	Continuous	Long
Decide Partnership	Continuous	Long
Decide Professional	Continuous	Long
Decide Standards	Continuous	Long
Exploit Consult	Continuous	Long
Exploit Electronic DB	Continuous	Long
Exploit References	Continuous	Long
Exploit Reflect	Continuous	Long
Exploit Simulate	Continuous	Long
IP	Key	Long
Learn BI	Continuous	Long
Learn Capture	Continuous	Long
Learn Mentor	Continuous	Long
Learn Offsite	Continuous	Long
Learn Portal	Continuous	Long
Learn Priority	Continuous	Long
Learn Reimburse	Continuous	Long
Learn Training	Continuous	Long
Learn Venue	Continuous	Long
Link Actively	Continuous	Long
Link Designated	Continuous	Long
Link Leadership	Continuous	Long
Link Monitor	Continuous	Long
Link Outsourcing	Continuous	Long
Link Relationship	Continuous	Long
OR Integer	Continuous	Long

Detect continuous or discrete for numeric columns:

Fig. A8.20: List of input variables used in model Cluster_Model1 – part 2



Create Testing Set

Specify the number of cases to be reserved for model testing.

Percentage of data for testing: 30 %

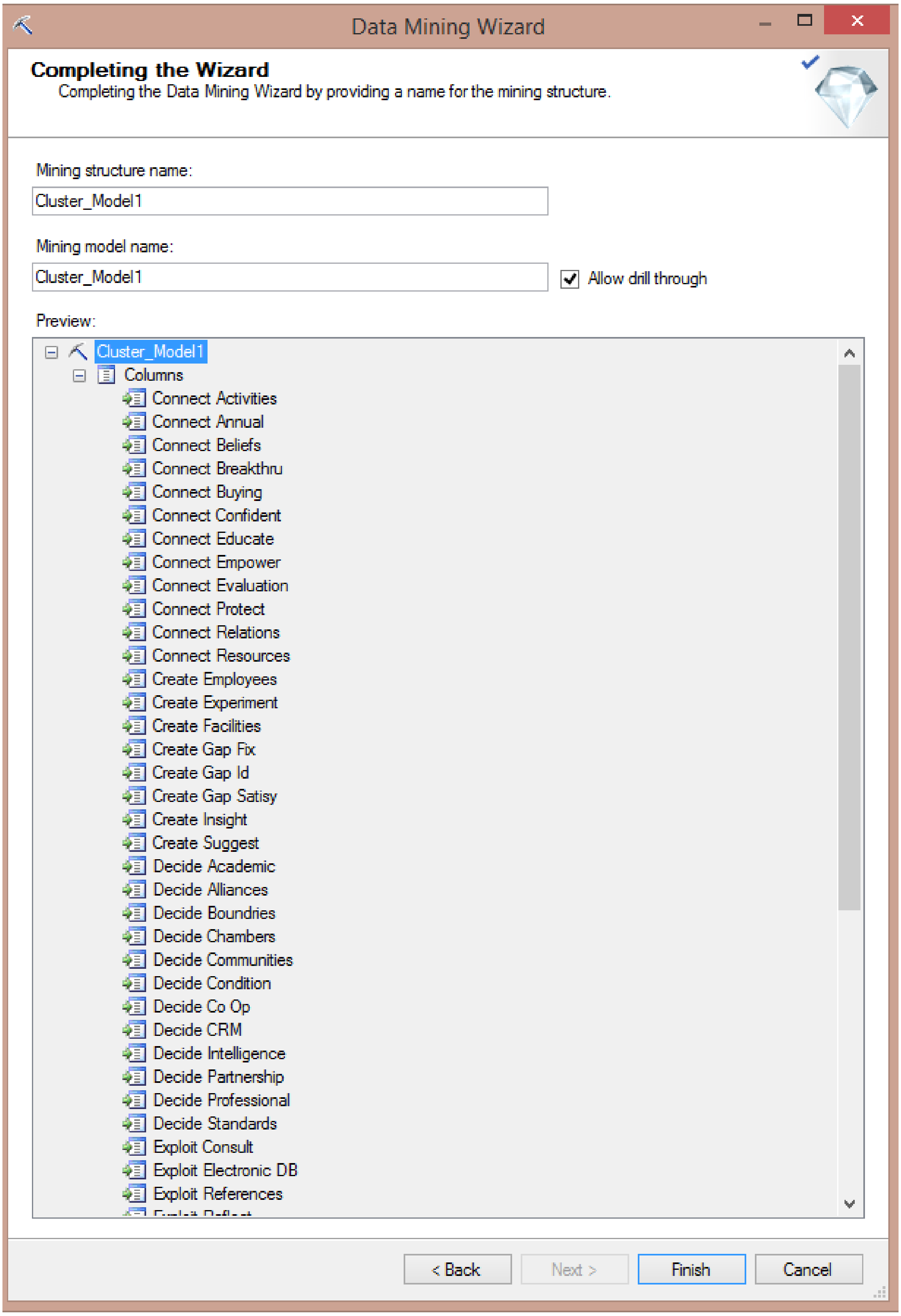
Maximum number of cases in testing data set:

Description:

Input data will be randomly split into two sets, a training set and a testing set, based on the percentage of data for testing and maximum number of cases in testing data set you provide. The training set is used to create the mining model. The testing set is used to check model accuracy.

[Percentage of data for testing] specifies percentages of cases reserved for testing set.
[Maximum number of cases in testing data set] limits total number of cases in the testing set.
If both values are specified, both limits are enforced.

Fig. A8.21: Allocation of data for testing



Data Mining Wizard

Completing the Wizard

Completing the Data Mining Wizard by providing a name for the mining structure.

Mining structure name:

Cluster_Model1

Mining model name:

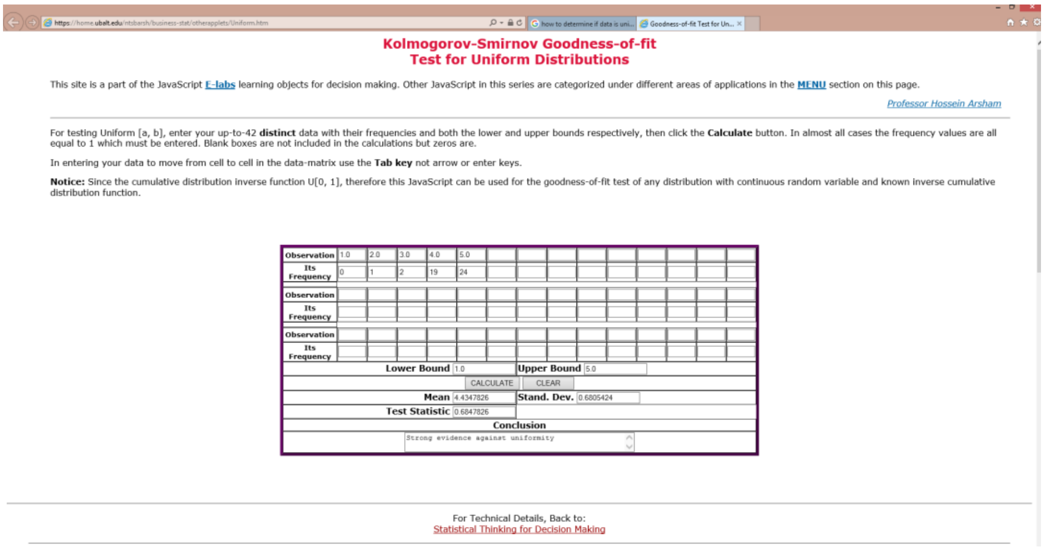
Cluster_Model1

☑ Allow drill through

Preview:

- Cluster_Model1
  - Columns
    - Connect Activities
    - Connect Annual
    - Connect Beliefs
    - Connect Breakthru
    - Connect Buying
    - Connect Confident
    - Connect Educate
    - Connect Empower
    - Connect Evaluation
    - Connect Protect
    - Connect Relations
    - Connect Resources
    - Create Employees
    - Create Experiment
    - Create Facilities
    - Create Gap Fix
    - Create Gap Id
    - Create Gap Satisy
    - Create Insight
    - Create Suggest
    - Decide Academic
    - Decide Alliances
    - Decide Boundries
    - Decide Chambers
    - Decide Communities
    - Decide Condition
    - Decide Co Op
    - Decide CRM
    - Decide Intelligence
    - Decide Partnership
    - Decide Professional
    - Decide Standards
    - Exploit Consult
    - Exploit Electronic DB
    - Exploit References
    - Exploit Reflect

Fig. A8.22: Completion of the DM wizard – ‘Allow drill through’ selected


https://home.ubalt.edu/ntsbarsh/business-stat/otherapplets/Uniform.htm
how to determine if data is uni...
Goodness-of-fit Test for Un...

Kolmogorov-Smirnov Goodness-of-fit
Test for Uniform Distributions

This site is a part of the JavaScript E-labs learning objects for decision making. Other JavaScript in this series are categorized under different areas of applications in the MENU section on this page.

Professor Hossein Arsham

For testing Uniform [a, b], enter your up-to-42 **distinct** data with their frequencies and both the lower and upper bounds respectively, then click the **Calculate** button. In almost all cases the frequency values are all equal to 1 which must be entered. Blank boxes are not included in the calculations but zeros are.

In entering your data to move from cell to cell in the data-matrix use the **Tab key** not arrow or enter keys.

Notice: Since the cumulative distribution inverse function U[0, 1], therefore this JavaScript can be used for the goodness-of-fit test of any distribution with continuous random variable and known inverse cumulative distribution function.

Observation	1.0	2.0	3.0	4.0	5.0									
Its Frequency	0	1	2	19	24									
Observation														
Its Frequency														
Observation														
Its Frequency														

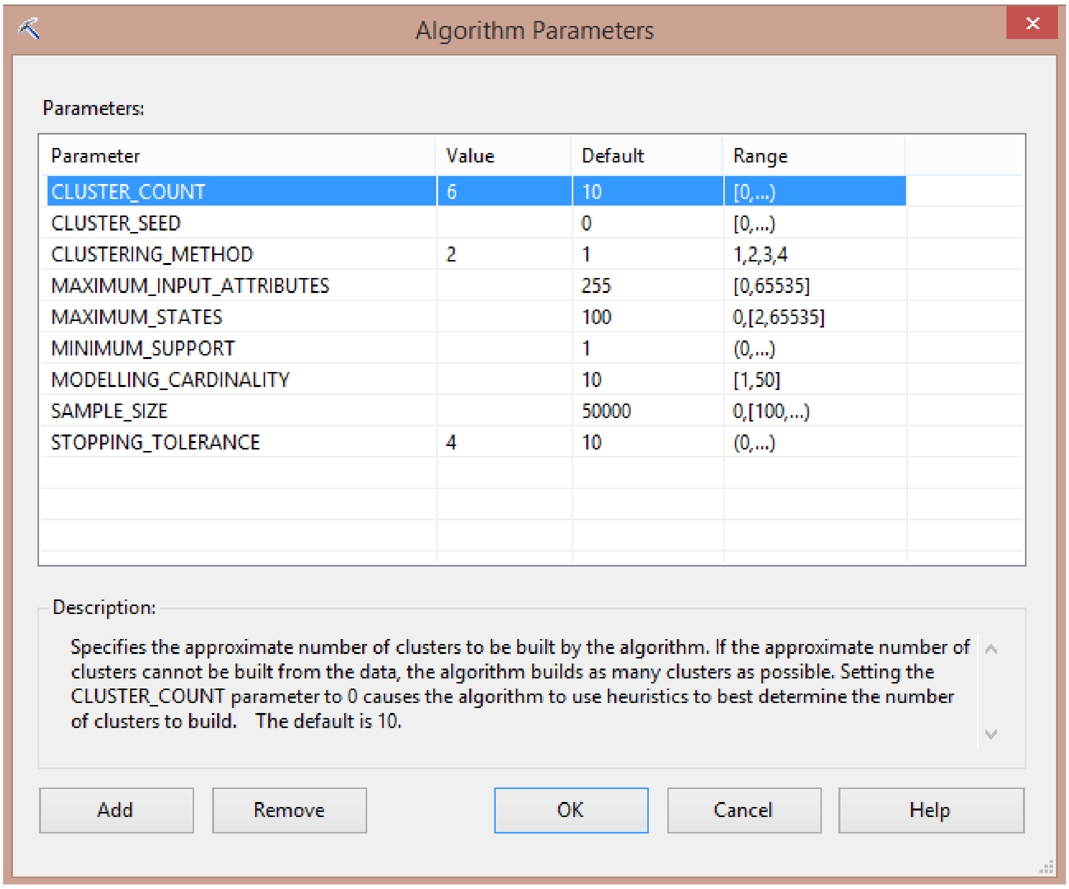
Lower Bound 1.0	Upper Bound 5.0
Mean 4.4347826	Stand. Dev. 0.6805424
Test Statistic 0.6847826	

Conclusion

Strong evidence against uniformity

For Technical Details, Back to:
Statistical Thinking for Decision Making

Fig. A8.23: The use of on-line tool (https://home.ubalt.edu/ntsbarsh/business-stat/otherapplets/Uniform.htm) to test for uniformly distributed data. Results are shown for the first question on the questionnaire: Create_GapID



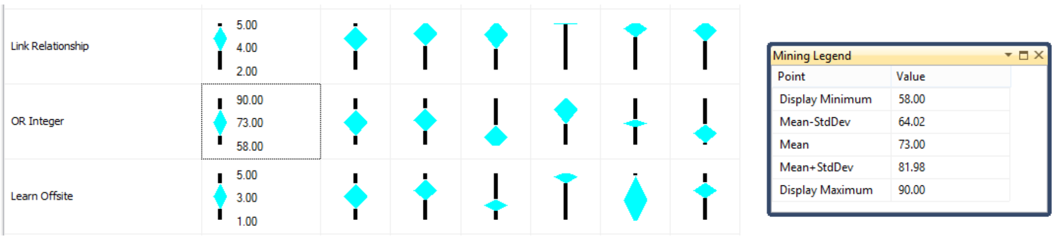
Parameters:

Parameter	Value	Default	Range
CLUSTER_COUNT	6	10	[0,...)
CLUSTER_SEED		0	[0,...)
CLUSTERING_METHOD	2	1	1,2,3,4
MAXIMUM_INPUT_ATTRIBUTES		255	[0,65535]
MAXIMUM_STATES		100	0,[2,65535]
MINIMUM_SUPPORT		1	(0,...)
MODELLING_CARDINALITY		10	[1,50]
SAMPLE_SIZE		50000	0,[100,...)
STOPPING_TOLERANCE	4	10	(0,...)

Description:

Specifies the approximate number of clusters to be built by the algorithm. If the approximate number of clusters cannot be built from the data, the algorithm builds as many clusters as possible. Setting the CLUSTER_COUNT parameter to 0 causes the algorithm to use heuristics to best determine the number of clusters to build. The default is 10.

Fig. A8.24: Clustering algorithm's parameters



Mining Legend

Point	Value
Display Minimum	58.00
Mean-StdDev	64.02
Mean	73.00
Mean+StdDev	81.98
Display Maximum	90.00

Fig. A8.25: Cluster profiles with the mining legend shown for the variable ‘OR Integer’

```
SELECT
  (Cluster()) as [ResultingCluster]
From
  [Cluster_Model1]
PREDICTION JOIN
  OPENQUERY([RGU_Analytics],
    'SELECT
      [Create_GapId],
      [Create_GapFix],
      [Create_GapSatisy],
      [Create_Employees],
      [Create_Facilities],
      [Create_Suggest],
      [Create_Experiment],
      [Create_Insight],
      [Exploit_References],
      [Exploit_Simulate],
      [Exploit_Consult],
      [Exploit_ElectronicDB],
      [Exploit_Reflect],
      [Decide_Alliances],
      [Decide_CoOp],
      [Decide_Partnership],
      [Decide_Standards],
      [Decide_professional],
      [Decide_Chambers],
      [Decide_Communities],
      [Decide_Academic],
      [Decide_Intelligence],
      [Decide_CRM],
      [Decide_Condition],
      [Decide_Boundries],
      [Learn_Training],
      [Learn_Mentor],
      [Learn_Reimburse],
      [Learn_Priority],
      [Learn_Capture],
      [Learn_Venue],
      [Learn_Offsite],
      [Learn_Portal],
      [Learn_BI],
      [Connect_Beliefs],
      [Connect_Empower],
      [Connect_Confident],
      [Connect_Breakthru],
      [Connect_Relations],
   [Connect_Evaluation],
   [Connect_Annual],
   [Connect_Educate],
   [Connect_Buying],
   [Connect_Activities],
   [Connect_Resources],
   [Connect_Protect],
   [Link_Relationship],
   [Link_Designated],
   [Link_Actively],
   [Link_Outsourcing],
   [Link_Monitor],
   [Link_Leadership]
  FROM
   [dbo].[tbl_NBModel2_Predict]
  ') AS t
ON
 [Cluster_Model1].[Create Gap Id] = t.[Create_GapId] AND
 [Cluster_Model1].[Create Gap Fix] = t.[Create_GapFix] AND
 [Cluster_Model1].[Create Gap Satisy] = t.[Create_GapSatisy] AND
 [Cluster_Model1].[Create Employees] = t.[Create_Employees] AND
 [Cluster_Model1].[Create Facilities] = t.[Create_Facilities] AND
 [Cluster_Model1].[Create Suggest] = t.[Create_Suggest] AND
 [Cluster_Model1].[Create Experiment] = t.[Create_Experiment] AND
 [Cluster_Model1].[Create Insight] = t.[Create_Insight] AND
 [Cluster_Model1].[Exploit References] = t.[Exploit_References] AND
 [Cluster_Model1].[Exploit Simulate] = t.[Exploit_Simulate] AND
 [Cluster_Model1].[Exploit Consult] = t.[Exploit_Consult] AND
 [Cluster_Model1].[Exploit Electronic DB] = t.[Exploit_ElectronicDB] AND
 [Cluster_Model1].[Exploit Reflect] = t.[Exploit_Reflect] AND
 [Cluster_Model1].[Decide Alliances] = t.[Decide_Alliances] AND
 [Cluster_Model1].[Decide Co Op] = t.[Decide_CoOp] AND
 [Cluster_Model1].[Decide Partnership] = t.[Decide_Partnership] AND
 [Cluster_Model1].[Decide Standards] = t.[Decide_Standards] AND
 [Cluster_Model1].[Decide Professional] = t.[Decide_professional] AND
 [Cluster_Model1].[Decide Chambers] = t.[Decide_Chambers] AND
 [Cluster_Model1].[Decide Communities] = t.[Decide_Communities] AND
 [Cluster_Model1].[Decide Academic] = t.[Decide_Academic] AND
 [Cluster_Model1].[Decide Intelligence] = t.[Decide_Intelligence] AND
 [Cluster_Model1].[Decide CRM] = t.[Decide_CRM] AND
 [Cluster_Model1].[Decide Condition] = t.[Decide_Condition] AND
 [Cluster_Model1].[Decide Boundries] = t.[Decide_Boundries] AND
 [Cluster_Model1].[Learn Training] = t.[Learn_Training] AND
 [Cluster_Model1].[Learn Mentor] = t.[Learn_Mentor] AND
 [Cluster_Model1].[Learn Reimburse] = t.[Learn_Reimburse] AND
 [Cluster_Model1].[Learn Priority] = t.[Learn_Priority] AND
 [Cluster_Model1].[Learn Capture] = t.[Learn_Capture] AND
 [Cluster_Model1].[Learn Venue] = t.[Learn_Venue] AND
 [Cluster_Model1].[Learn Offsite] = t.[Learn_Offsite] AND
 [Cluster_Model1].[Learn Portal] = t.[Learn_Portal] AND
 [Cluster_Model1].[Learn BI] = t.[Learn_BI] AND
 [Cluster_Model1].[Connect Beliefs] = t.[Connect_Beliefs] AND
 [Cluster_Model1].[Connect Empower] = t.[Connect_Empower] AND
 [Cluster_Model1].[Connect Confident] = t.[Connect_Confident] AND
 [Cluster_Model1].[Connect Breakthru] = t.[Connect_Breakthru] AND
 [Cluster_Model1].[Connect Relations] = t.[Connect_Relations] AND
 [Cluster_Model1].[Connect Evaluation] = t.[Connect_Evaluation] AND
 [Cluster_Model1].[Connect Annual] = t.[Connect_Annual] AND
 [Cluster_Model1].[Connect Educate] = t.[Connect_Educate] AND
 [Cluster_Model1].[Connect Buying] = t.[Connect_Buying] AND
 [Cluster_Model1].[Connect Activities] = t.[Connect_Activities] AND
 [Cluster_Model1].[Connect Resources] = t.[Connect_Resources] AND
 [Cluster_Model1].[Connect Protect] = t.[Connect_Protect] AND
 [Cluster_Model1].[Link Relationship] = t.[Link_Relationship] AND
 [Cluster_Model1].[Link Designated] = t.[Link_Designated] AND
```

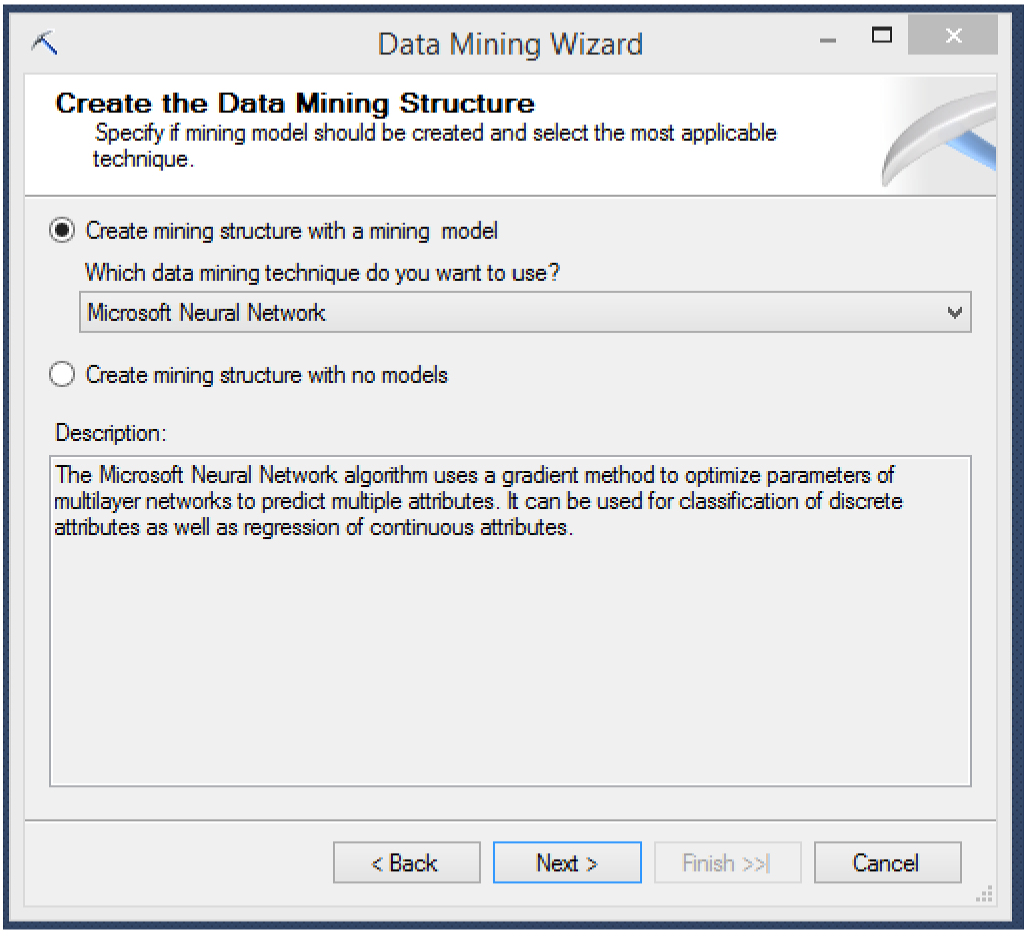
```
  [Cluster_Model1].[Link Actively] = t.[Link_Actively] AND
  [Cluster_Model1].[Link Outsourcing] = t.[Link_Outsourcing] AND
  [Cluster_Model1].[Link Monitor] = t.[Link_Monitor] AND
  [Cluster_Model1].[Link Leadership] = t.[Link_Leadership]
```

Fig. A8.26: Predictive query used to obtain the cluster number

```
SELECT
  ([Cluster_Model1].[OR Integer]) as [Model Data OR Score],
  (t.[ORInteger]) as [Input Table Data OR Score]
From
  [Cluster_Model1]
PREDICTION JOIN
  OPENQUERY([RGU_Analytics],
    'SELECT
      [ORInteger],
      [Create_GapId],
      [Create_GapFix],
      [Create_GapSatisy],
      [Create_Employees],
      [Create_Facilities],
      [Create_Suggest],
      [Create_Experiment],
      [Create_Insight],
      [Exploit_References],
      [Exploit_Simulate],
      [Exploit_Consult],
      [Exploit_ElectronicDB],
      [Exploit_Reflect],
      [Decide_Alliances],
      [Decide_CoOp],
      [Decide_Partnership],
      [Decide_Standards],
      [Decide_professional],
      [Decide_Chambers],
      [Decide_Communities],
      [Decide_Academic],
      [Decide_Intelligence],
      [Decide_CRM],
      [Decide_Condition],
      [Decide_Boundries],
      [Learn_Training],
      [Learn_Mentor],
      [Learn_Reimburse],
      [Learn_Priority],
      [Learn_Capture],
      [Learn_Venue],
      [Learn_Offsite],
      [Learn_Portal],
      [Learn_BI],
      [Connect_Beliefs],
      [Connect_Empower],
      [Connect_Confident],
      [Connect_Breakthru],
      [Connect_Relations],
      [Connect_Evaluation],
      [Connect_Annual],
      [Connect_Educate],
      [Connect_Buying],
      [Connect_Activities],
      [Connect_Resources],
      [Connect_Protect],
      [Link_Relationship],
      [Link_Designated],
      [Link_Actively],
      [Link_Outsourcing],
      [Link_Monitor],
      [Link_Leadership]
    FROM
      [dbo].[tbl_NBModel2_Predict]
    ') AS t
ON
  [Cluster_Model1].[Create Gap Id] = t.[Create_GapId] AND
  [Cluster_Model1].[Create Gap Fix] = t.[Create_GapFix] AND
  [Cluster_Model1].[Create Gap Satisy] = t.[Create_GapSatisy] AND
  [Cluster_Model1].[Create Employees] = t.[Create_Employees] AND
  [Cluster_Model1].[Create Facilities] = t.[Create_Facilities] AND
  [Cluster_Model1].[Create Suggest] = t.[Create_Suggest] AND
  [Cluster_Model1].[Create Experiment] = t.[Create_Experiment] AND
  [Cluster_Model1].[Create Insight] = t.[Create_Insight] AND
  [Cluster_Model1].[Exploit References] = t.[Exploit_References] AND
  [Cluster_Model1].[Exploit Simulate] = t.[Exploit_Simulate] AND
  [Cluster_Model1].[Exploit Consult] = t.[Exploit_Consult] AND
  [Cluster_Model1].[Exploit Electronic DB] = t.[Exploit_ElectronicDB] AND
  [Cluster_Model1].[Exploit Reflect] = t.[Exploit_Reflect] AND
  [Cluster_Model1].[Decide Alliances] = t.[Decide_Alliances] AND
  [Cluster_Model1].[Decide Co Op] = t.[Decide_CoOp] AND
  [Cluster_Model1].[Decide Partnership] = t.[Decide_Partnership] AND
  [Cluster_Model1].[Decide Standards] = t.[Decide_Standards] AND
  [Cluster_Model1].[Decide Professional] = t.[Decide_professional] AND
  [Cluster_Model1].[Decide Chambers] = t.[Decide_Chambers] AND
  [Cluster_Model1].[Decide Communities] = t.[Decide_Communities] AND
  [Cluster_Model1].[Decide Academic] = t.[Decide_Academic] AND
  [Cluster_Model1].[Decide Intelligence] = t.[Decide_Intelligence] AND
  [Cluster_Model1].[Decide CRM] = t.[Decide_CRM] AND
  [Cluster_Model1].[Decide Condition] = t.[Decide_Condition] AND
  [Cluster_Model1].[Decide Boundries] = t.[Decide_Boundries] AND
  [Cluster_Model1].[Learn Training] = t.[Learn_Training] AND
  [Cluster_Model1].[Learn Mentor] = t.[Learn_Mentor] AND
  [Cluster_Model1].[Learn Reimburse] = t.[Learn_Reimburse] AND
  [Cluster_Model1].[Learn Priority] = t.[Learn_Priority] AND
  [Cluster_Model1].[Learn Capture] = t.[Learn_Capture] AND
  [Cluster_Model1].[Learn Venue] = t.[Learn_Venue] AND
  [Cluster_Model1].[Learn Offsite] = t.[Learn_Offsite] AND
  [Cluster_Model1].[Learn Portal] = t.[Learn_Portal] AND
  [Cluster_Model1].[Learn BI] = t.[Learn_BI] AND
  [Cluster_Model1].[Connect Beliefs] = t.[Connect_Beliefs] AND
  [Cluster_Model1].[Connect Empower] = t.[Connect_Empower] AND
  [Cluster_Model1].[Connect Confident] = t.[Connect_Confident] AND
  [Cluster_Model1].[Connect Breakthru] = t.[Connect_Breakthru] AND
  [Cluster_Model1].[Connect Relations] = t.[Connect_Relations] AND
  [Cluster_Model1].[Connect Evaluation] = t.[Connect_Evaluation] AND
  [Cluster_Model1].[Connect Annual] = t.[Connect_Annual] AND
  [Cluster_Model1].[Connect Educate] = t.[Connect_Educate] AND
  [Cluster_Model1].[Connect Buying] = t.[Connect_Buying] AND
  [Cluster_Model1].[Connect Activities] = t.[Connect_Activities] AND
  [Cluster_Model1].[Connect Resources] = t.[Connect_Resources] AND
  [Cluster_Model1].[Connect Protect] = t.[Connect_Protect] AND
  [Cluster_Model1].[Link Relationship] = t.[Link_Relationship] AND
  [Cluster_Model1].[Link Designated] = t.[Link_Designated] AND
  [Cluster_Model1].[Link Actively] = t.[Link_Actively] AND
```

```
[Cluster_Model1].[Link Outsourcing] = t.[Link_Outsourcing] AND
[Cluster_Model1].[Link Monitor] = t.[Link_Monitor] AND
[Cluster_Model1].[Link Leadership] = t.[Link_Leadership]
```

Fig. A8.27: Predictive query used to obtain 'OR Integer' value



Data Mining Wizard

Create the Data Mining Structure

Specify if mining model should be created and select the most applicable technique.

● Create mining structure with a mining model

Which data mining technique do you want to use?

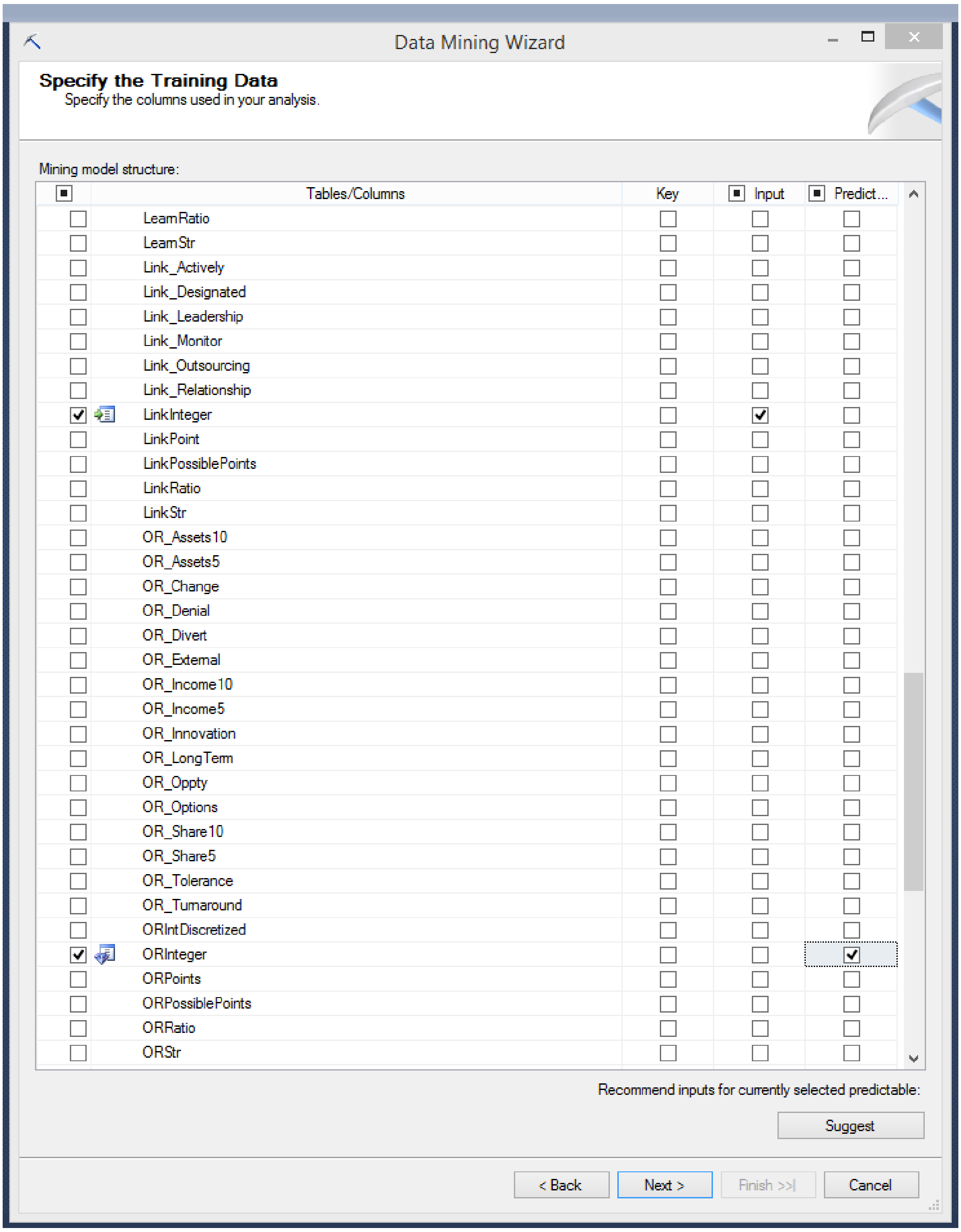
Microsoft Neural Network

○ Create mining structure with no models

Description:

The Microsoft Neural Network algorithm uses a gradient method to optimize parameters of multilayer networks to predict multiple attributes. It can be used for classification of discrete attributes as well as regression of continuous attributes.

Fig. A8.28 Construction of the neural network model: NN_Model1



Data Mining Wizard

Specify the Training Data

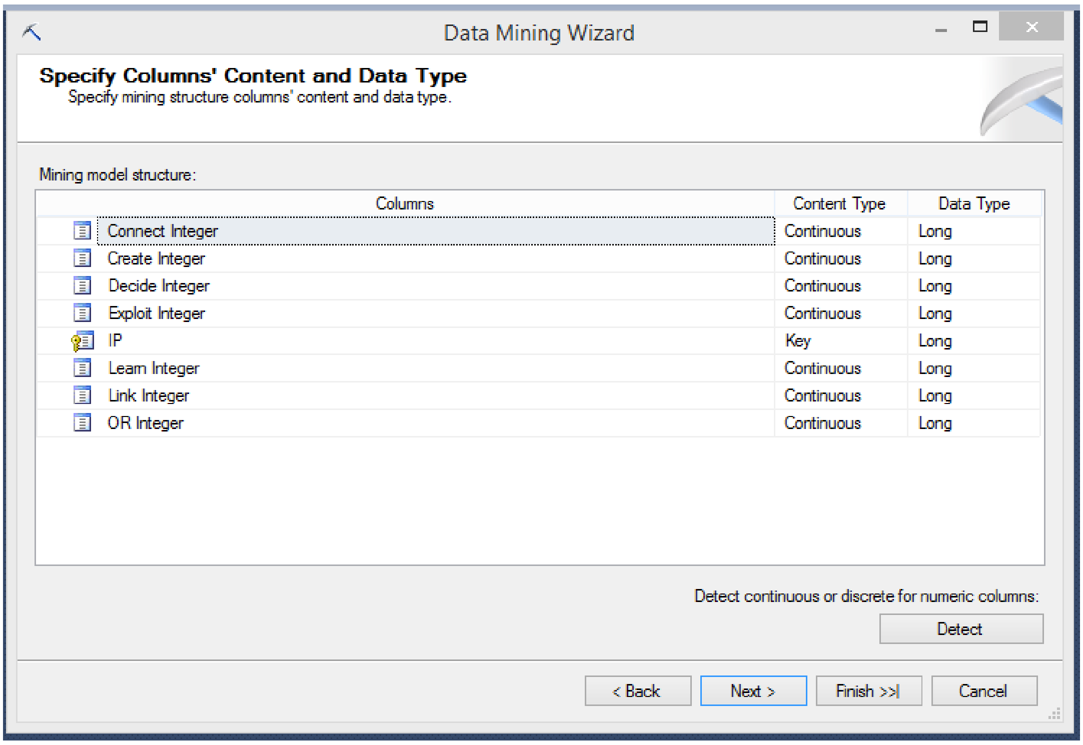
Specify the columns used in your analysis.

Mining model structure:

■	Tables/Columns	Key	■ Input	■ Predict...
□	LearnRatio	□	□	□
□	LearnStr	□	□	□
□	Link_Actively	□	□	□
□	Link_Designated	□	□	□
□	Link_Leadership	□	□	□
□	Link_Monitor	□	□	□
□	Link_Outsourcing	□	□	□
□	Link_Relationship	□	□	□
☑	LinkInteger	□	☑	□
□	LinkPoint	□	□	□
□	LinkPossiblePoints	□	□	□
□	LinkRatio	□	□	□
□	LinkStr	□	□	□
□	OR_Assets10	□	□	□
□	OR_Assets5	□	□	□
□	OR_Change	□	□	□
□	OR_Denial	□	□	□
□	OR_Divert	□	□	□
□	OR_External	□	□	□
□	OR_Income10	□	□	□
□	OR_Income5	□	□	□
□	OR_Innovation	□	□	□
□	OR_LongTerm	□	□	□
□	OR_Oppty	□	□	□
□	OR_Options	□	□	□
□	OR_Share10	□	□	□
□	OR_Share5	□	□	□
□	OR_Tolerance	□	□	□
□	OR_Turnaround	□	□	□
□	ORIntDiscretized	□	□	□
☑	ORInteger	□	□	☑
□	ORPoints	□	□	□
□	ORPossiblePoints	□	□	□
□	ORRatio	□	□	□
□	ORStr	□	□	□

Recommend inputs for currently selected predictable:

Fig. A8.29: Neural network model construction – specification of columns used in analysis



Specify Columns' Content and Data Type

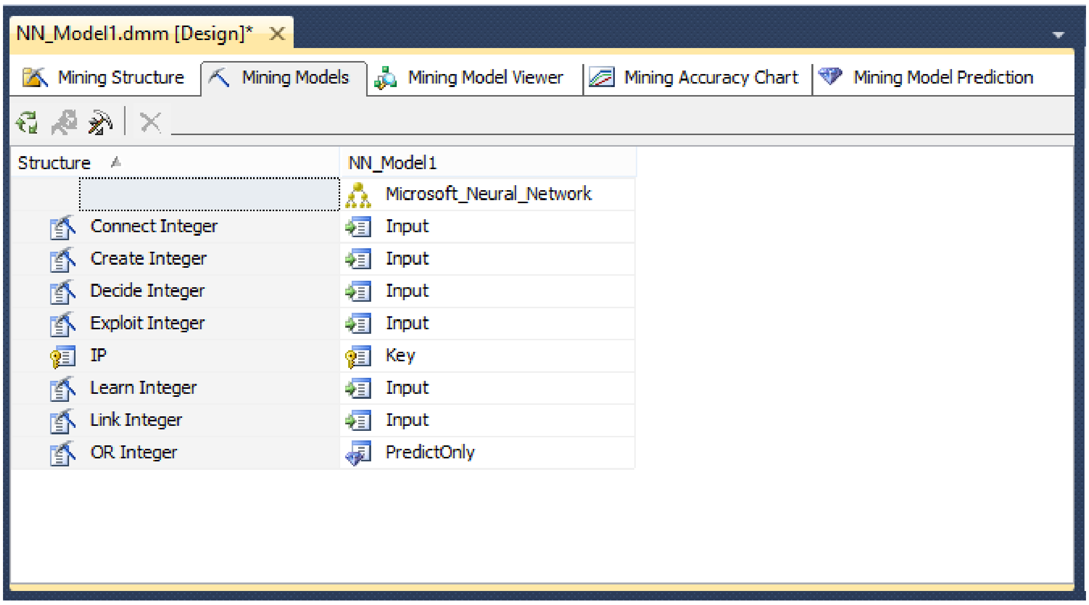
Specify mining structure columns' content and data type.

Mining model structure:

Columns	Content Type	Data Type
Connect Integer	Continuous	Long
Create Integer	Continuous	Long
Decide Integer	Continuous	Long
Exploit Integer	Continuous	Long
IP	Key	Long
Learn Integer	Continuous	Long
Link Integer	Continuous	Long
OR Integer	Continuous	Long

Detect continuous or discrete for numeric columns:

Fig. A8.30: Data types used in construction of neural network model

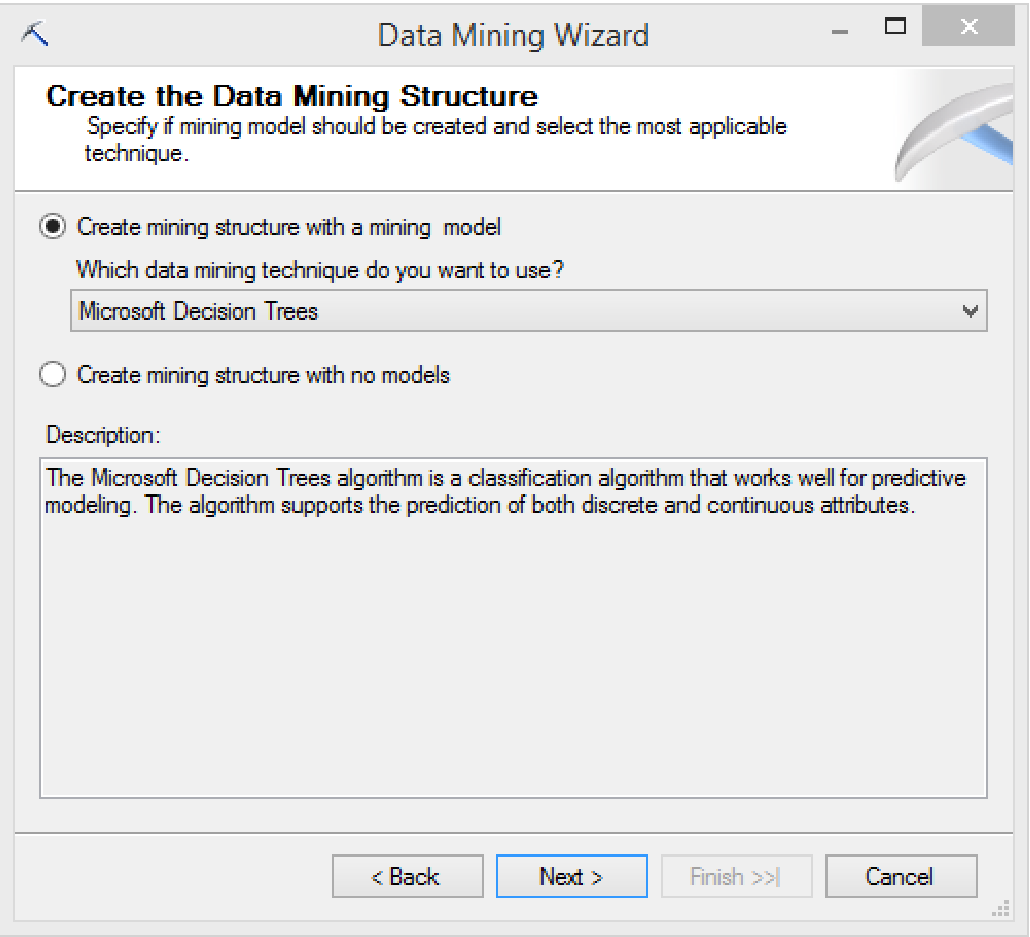


Structure	NN_Model1
	Microsoft_Neural_Network
Connect Integer	Input
Create Integer	Input
Decide Integer	Input
Exploit Integer	Input
IP	Key
Learn Integer	Input
Link Integer	Input
OR Integer	PredictOnly

Fig. A.8.31: The NN mining model: NN_Model1

```
SELECT
  [NN_Model1].[OR Integer]
From
  [NN_Model1]
NATURAL PREDICTION JOIN
(SELECT 70 AS [Connect Integer],
70 AS [Create Integer],
67 AS [Decide Integer],
56 AS [Exploit Integer],
67 AS [Learn Integer],
73 AS [Link Integer]) AS t
```

Fig. A.8.32 Query used in the output from NN model: NN_Model1



Create the Data Mining Structure

Specify if mining model should be created and select the most applicable technique.

● Create mining structure with a mining model

Which data mining technique do you want to use?

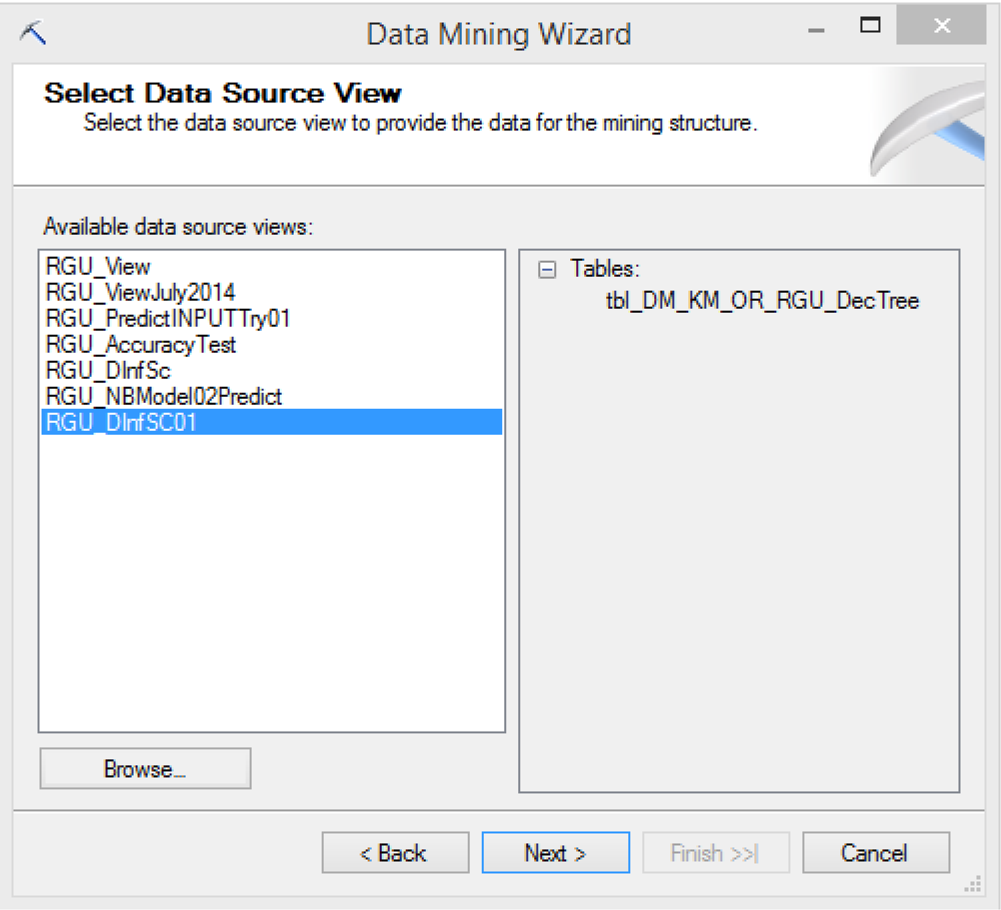
Microsoft Decision Trees

○ Create mining structure with no models

Description:

The Microsoft Decision Trees algorithm is a classification algorithm that works well for predictive modeling. The algorithm supports the prediction of both discrete and continuous attributes.

Fig. A8.33: DT Construction, initial step



Data Mining Wizard

Select Data Source View

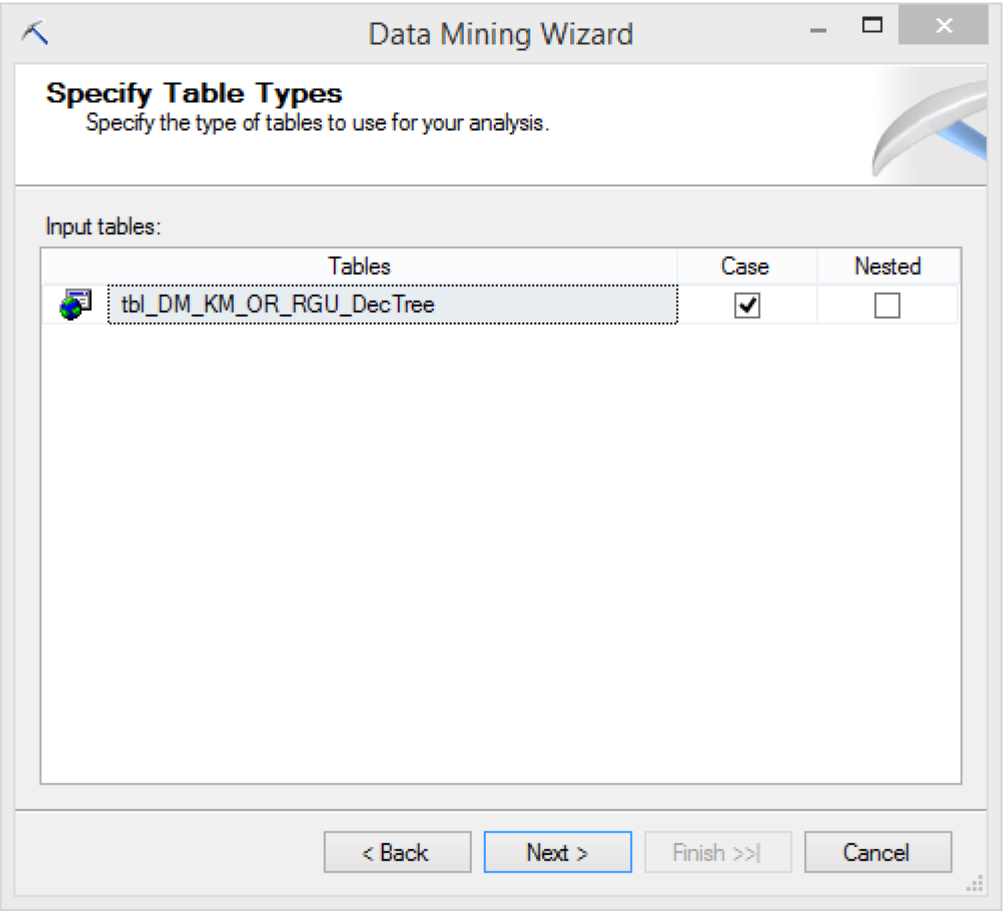
Select the data source view to provide the data for the mining structure.

Available data source views:

RGU_View
RGU_ViewJuly2014
RGU_PredictINPUTTry01
RGU_AccuracyTest
RGU_DInfSc
RGU_NBModel02Predict
RGU_DInfSC01

⊟ Tables:
tbl_DM_KM_OR_RGU_DecTree

Fig. A8.34: DT model construction: creation of data source view



Data Mining Wizard

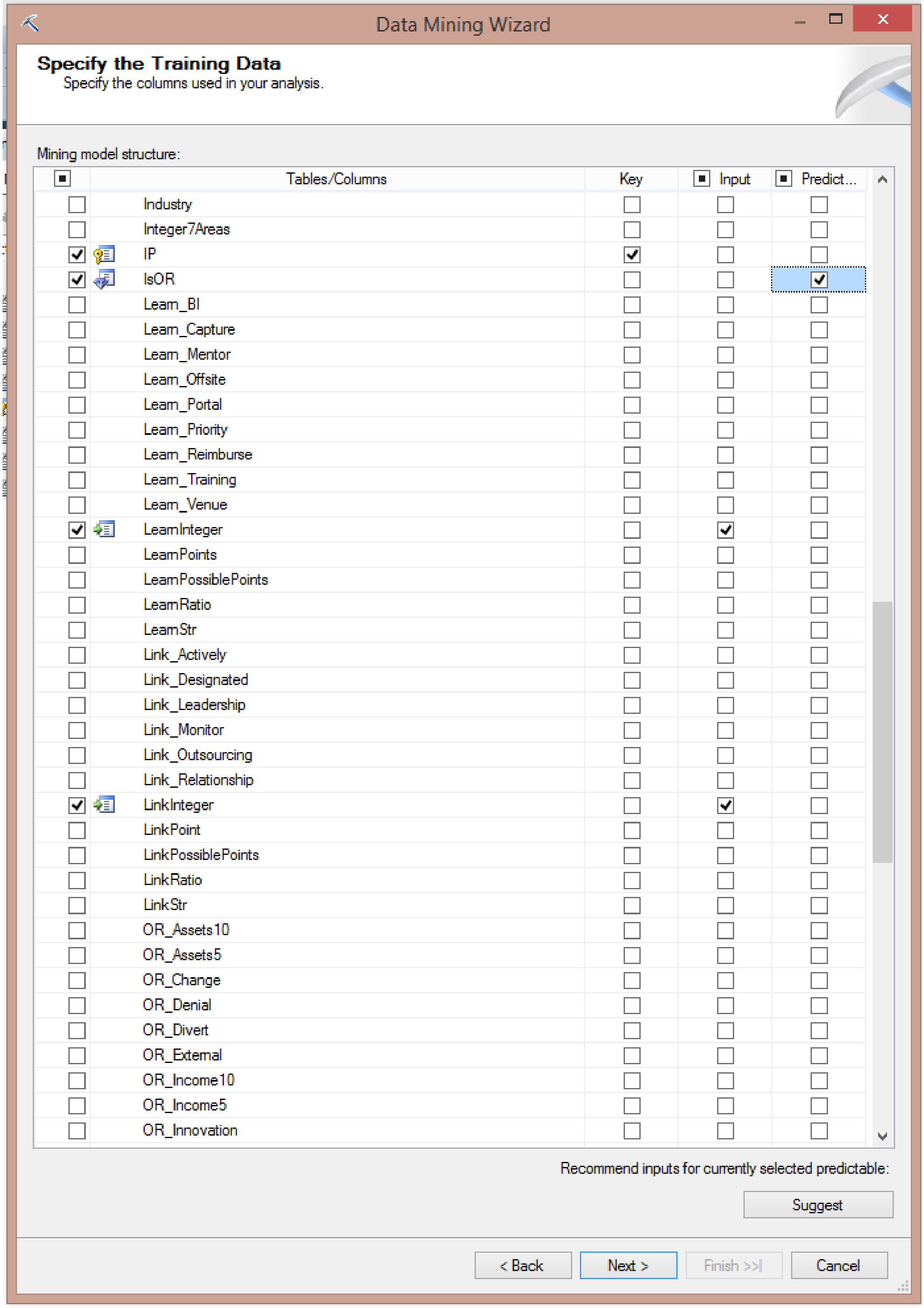
Specify Table Types

Specify the type of tables to use for your analysis.

Input tables:

Tables	Case	Nested
tbl_DM_KM_OR_RGU_DecTree	☑	□

Fig A.8.35: Construction of the DT model using DM wizard



Data Mining Wizard

Specify the Training Data

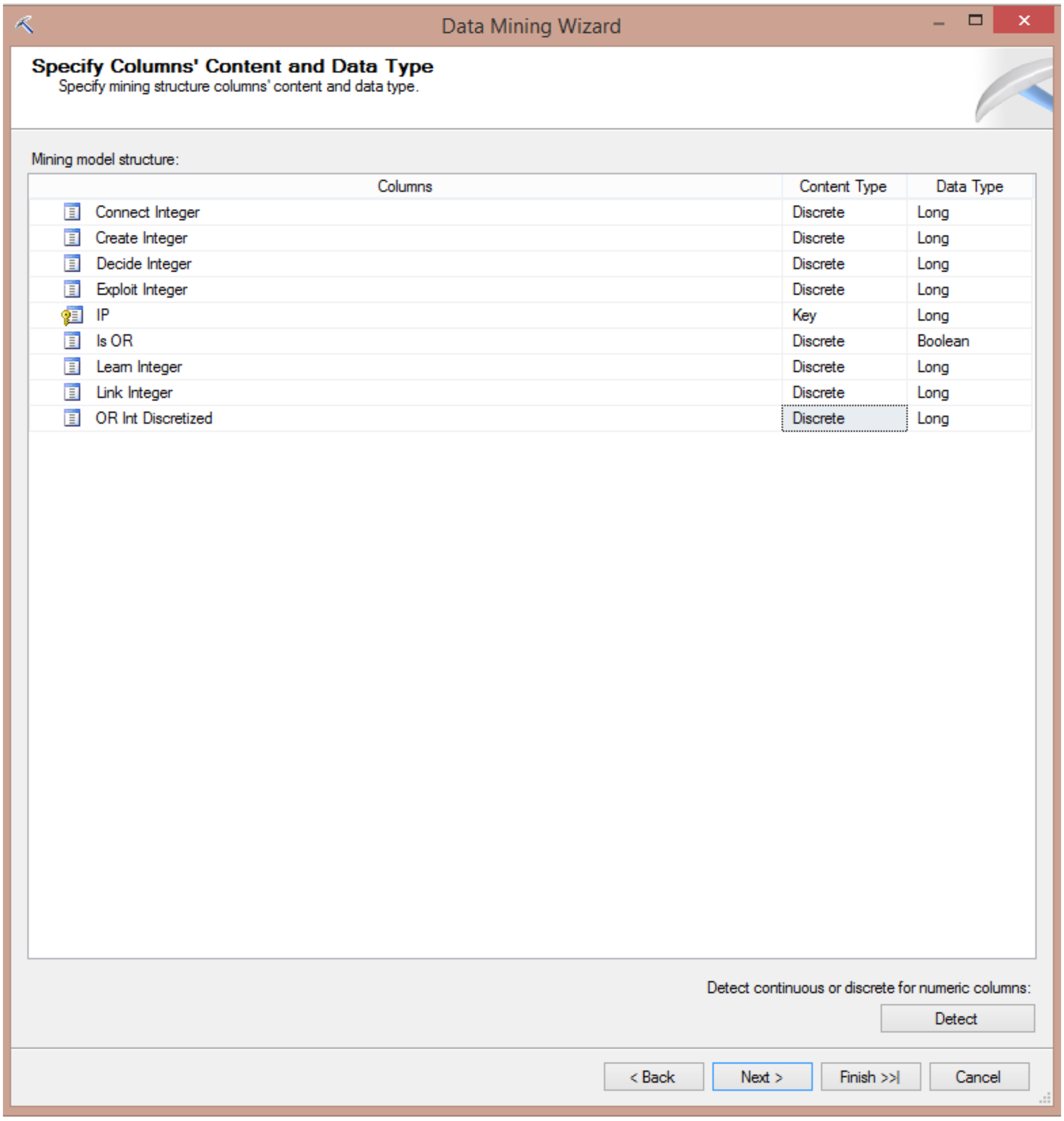
Specify the columns used in your analysis.

Mining model structure:

■	Tables/Columns	Key	■ Input	■ Predict...
□	Industry	□	□	□
□	Integer7Areas	□	□	□
☑	IP	☑	□	□
☑	IsOR	□	□	☑
□	Learn_BI	□	□	□
□	Learn_Capture	□	□	□
□	Learn_Mentor	□	□	□
□	Learn_Offsite	□	□	□
□	Learn_Portal	□	□	□
□	Learn_Priority	□	□	□
□	Learn_Reimburse	□	□	□
□	Learn_Training	□	□	□
□	Learn_Venue	□	□	□
☑	LearnInteger	□	☑	□
□	LearnPoints	□	□	□
□	LearnPossiblePoints	□	□	□
□	LearnRatio	□	□	□
□	LearnStr	□	□	□
□	Link_Actively	□	□	□
□	Link_Designated	□	□	□
□	Link_Leadership	□	□	□
□	Link_Monitor	□	□	□
□	Link_Outsourcing	□	□	□
□	Link_Relationship	□	□	□
☑	LinkInteger	□	☑	□
□	LinkPoint	□	□	□
□	LinkPossiblePoints	□	□	□
□	LinkRatio	□	□	□
□	LinkStr	□	□	□
□	OR_Assets10	□	□	□
□	OR_Assets5	□	□	□
□	OR_Change	□	□	□
□	OR_Denial	□	□	□
□	OR_Divert	□	□	□
□	OR_External	□	□	□
□	OR_Income10	□	□	□
□	OR_Income5	□	□	□
□	OR_Innovation	□	□	□

Recommend inputs for currently selected predictable:

Fig. A8.36: DT model construction, selection of training data



Specify Columns' Content and Data Type

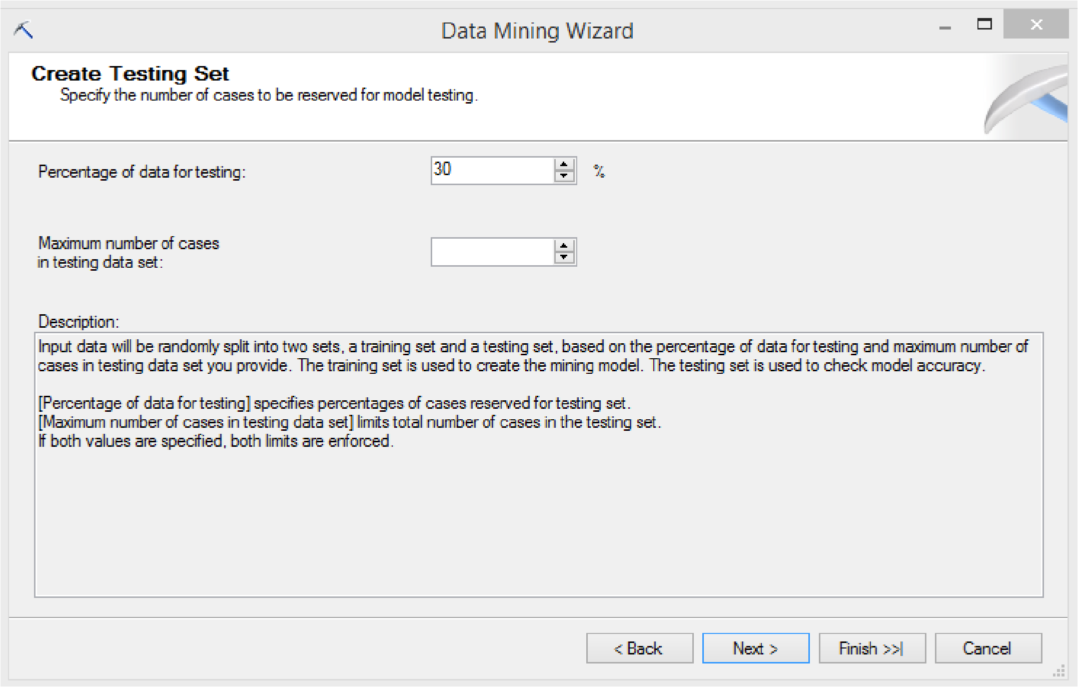
Specify mining structure columns' content and data type.

Mining model structure:

Columns	Content Type	Data Type
Connect Integer	Discrete	Long
Create Integer	Discrete	Long
Decide Integer	Discrete	Long
Exploit Integer	Discrete	Long
IP	Key	Long
Is OR	Discrete	Boolean
Learn Integer	Discrete	Long
Link Integer	Discrete	Long
OR Int Discretized	Discrete	Long

Detect continuous or discrete for numeric columns:

Fig. A8.37: DT Model’s mining structure specification



Create Testing Set

Specify the number of cases to be reserved for model testing.

Percentage of data for testing: 30 %

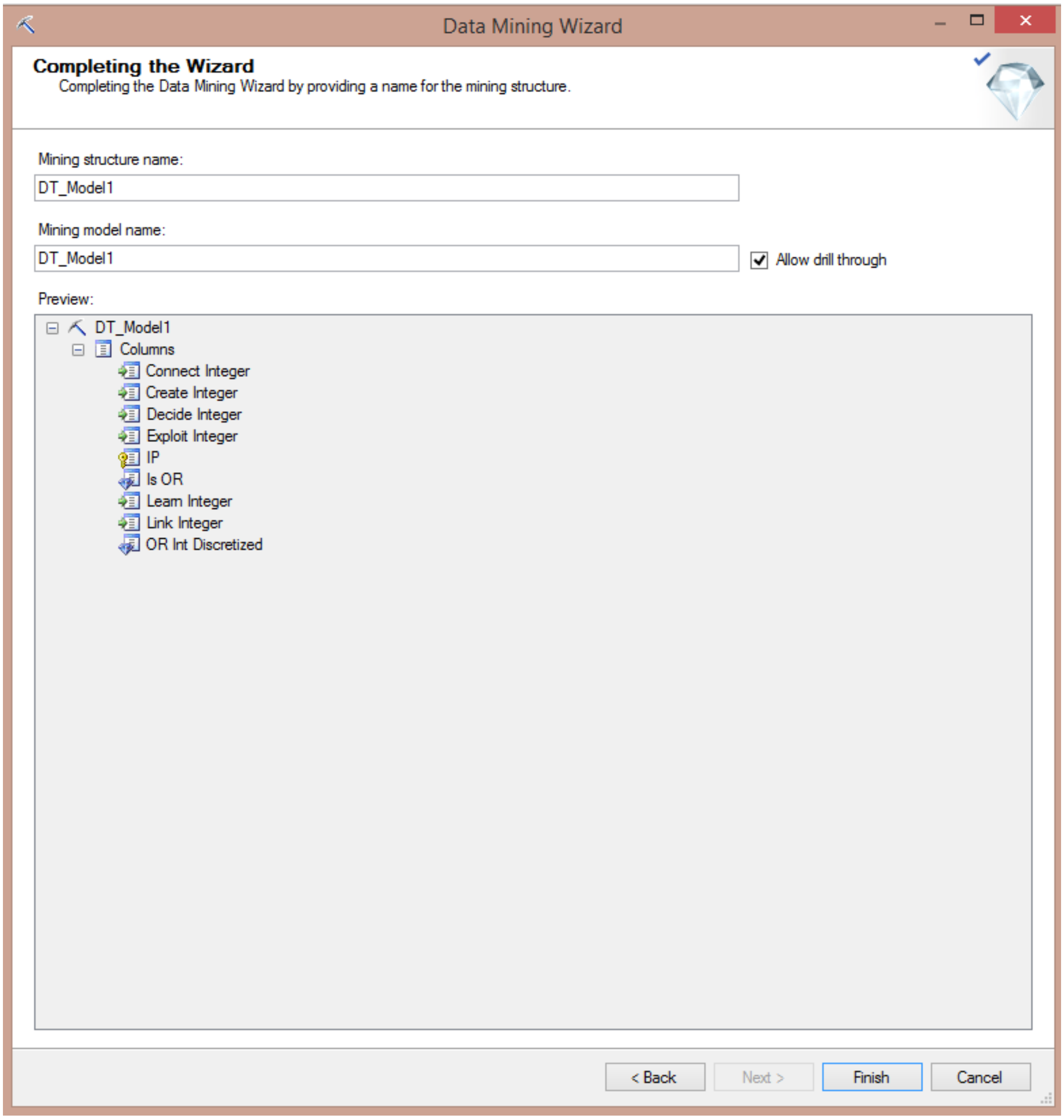
Maximum number of cases in testing data set:

Description:

Input data will be randomly split into two sets, a training set and a testing set, based on the percentage of data for testing and maximum number of cases in testing data set you provide. The training set is used to create the mining model. The testing set is used to check model accuracy.

[Percentage of data for testing] specifies percentages of cases reserved for testing set.
[Maximum number of cases in testing data set] limits total number of cases in the testing set.
If both values are specified, both limits are enforced.

Fig. A8.38: DT model's setting aside testing cases



Completing the Wizard

Completing the Data Mining Wizard by providing a name for the mining structure.

Mining structure name:

DT_Model1

Mining model name:

DT_Model1

☑ Allow drill through

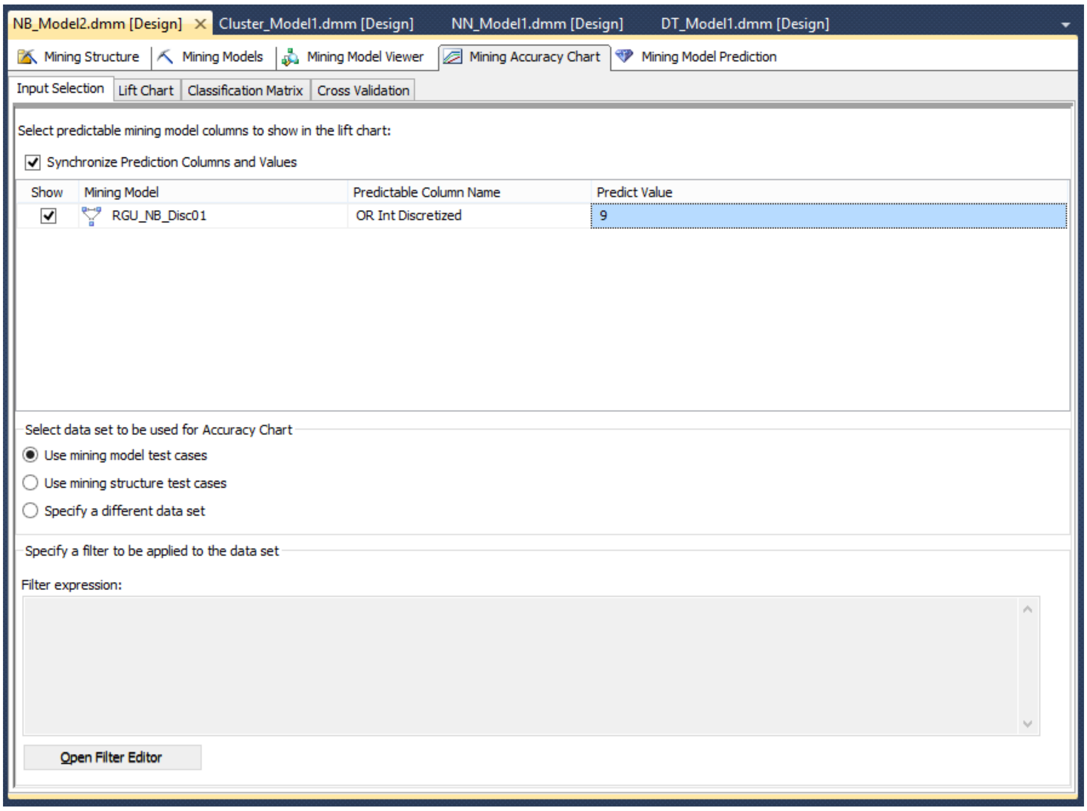
Preview:

- DT_Model1
  - Columns
    - Connect Integer
    - Create Integer
    - Decide Integer
    - Exploit Integer
    - IP
    - Is OR
    - Learn Integer
    - Link Integer
    - OR Int Discretized

Fig. A8.39: Final step in DT model construction: naming the model as 'drill through' option selection

```
SELECT
  [DT_Model1].[Is OR],
  [DT_Model1].[OR Int Discretized]
From
  [DT_Model1]
NATURAL PREDICTION JOIN
(SELECT 70 AS [Connect Integer],
  70 AS [Create Integer],
  67 AS [Decide Integer],
  56 AS [Exploit Integer],
  67 AS [Learn Integer],
  73 AS [Link Integer]) AS t
```

Fig. A8.40: DT prediction model's query



Select predictable mining model columns to show in the lift chart:

☑ Synchronize Prediction Columns and Values

Show	Mining Model	Predictable Column Name	Predict Value
☑	RGU_NB_Disc01	OR Int Discretized	9

Select data set to be used for Accuracy Chart

● Use mining model test cases

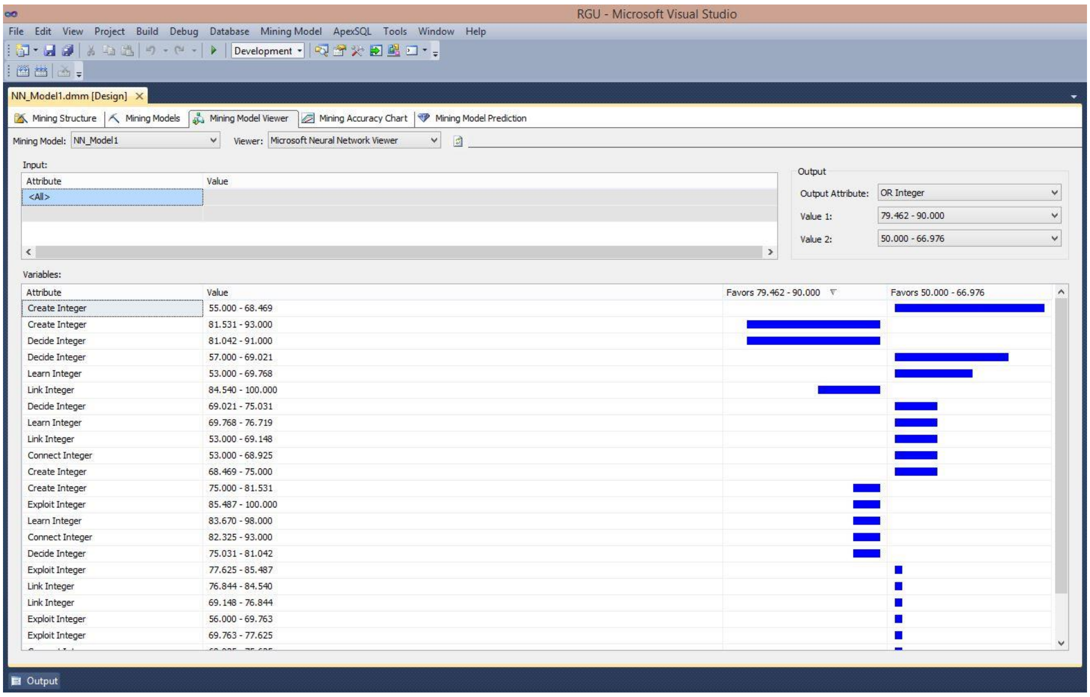
○ Use mining structure test cases

○ Specify a different data set

Specify a filter to be applied to the data set

Filter expression:

Fig. A8.41: Example of selection of input mining column and predictable value



Mining Model: NN_Model1 Viewer: Microsoft Neural Network Viewer

Input:

Attribute	Value
<All>	

Output

Output Attribute: OR Integer

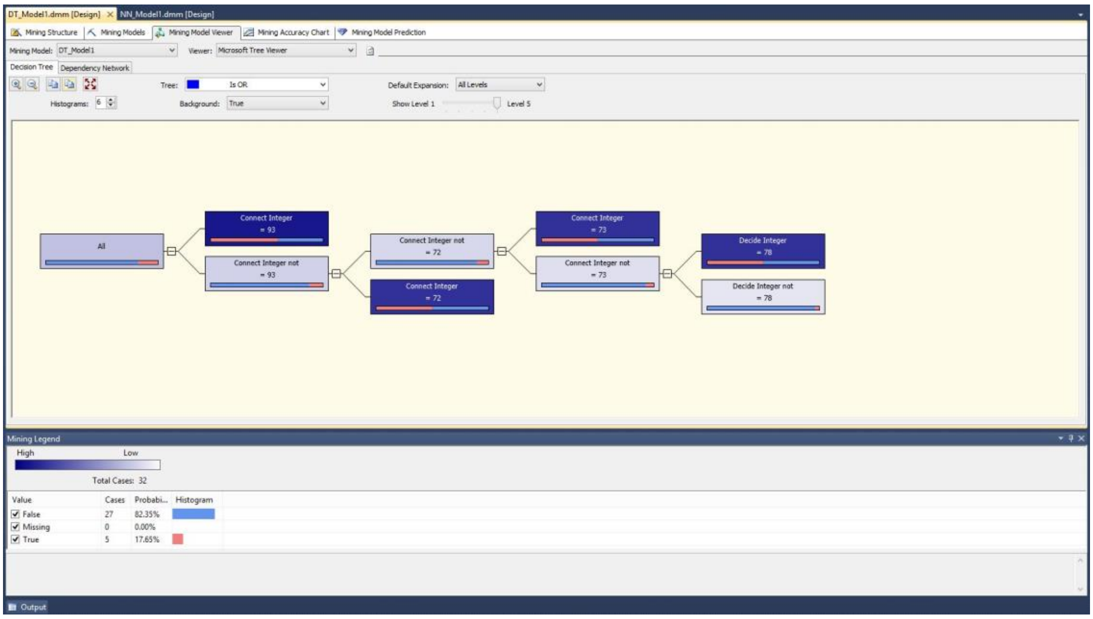
Value 1: 79.462 - 90.000

Value 2: 50.000 - 66.976

Variables:

Attribute	Value	Favors 79.462 - 90.000	Favors 50.000 - 66.976
Create Integer	55.000 - 68.469		■
Create Integer	81.531 - 93.000	■	
Decide Integer	81.042 - 91.000	■	
Decide Integer	57.000 - 69.021		■
Learn Integer	53.000 - 69.768		■
Link Integer	84.540 - 100.000	■	
Decide Integer	69.021 - 75.031		■
Learn Integer	69.768 - 76.719		■
Link Integer	53.000 - 69.148		■
Connect Integer	53.000 - 68.925		■
Create Integer	68.469 - 75.000		■
Create Integer	75.000 - 81.531	■	
Exploit Integer	85.487 - 100.000	■	
Learn Integer	83.670 - 98.000	■	
Connect Integer	82.325 - 93.000	■	
Decide Integer	75.031 - 81.042	■	
Exploit Integer	77.625 - 85.487		■
Link Integer	76.844 - 84.540		■
Link Integer	69.148 - 76.844		■
Exploit Integer	56.000 - 69.763		■
Exploit Integer	69.763 - 77.625		■

Fig. A8.42: Example of the NN model showing key competence areas and their values with respect to the highest and lowest 'OR Score'



Mining Model: DT_Model1

Viewer: Microsoft Tree Viewer

Tree: Is OR

Default Expansion: All Levels

Histograms: 6

Background: True

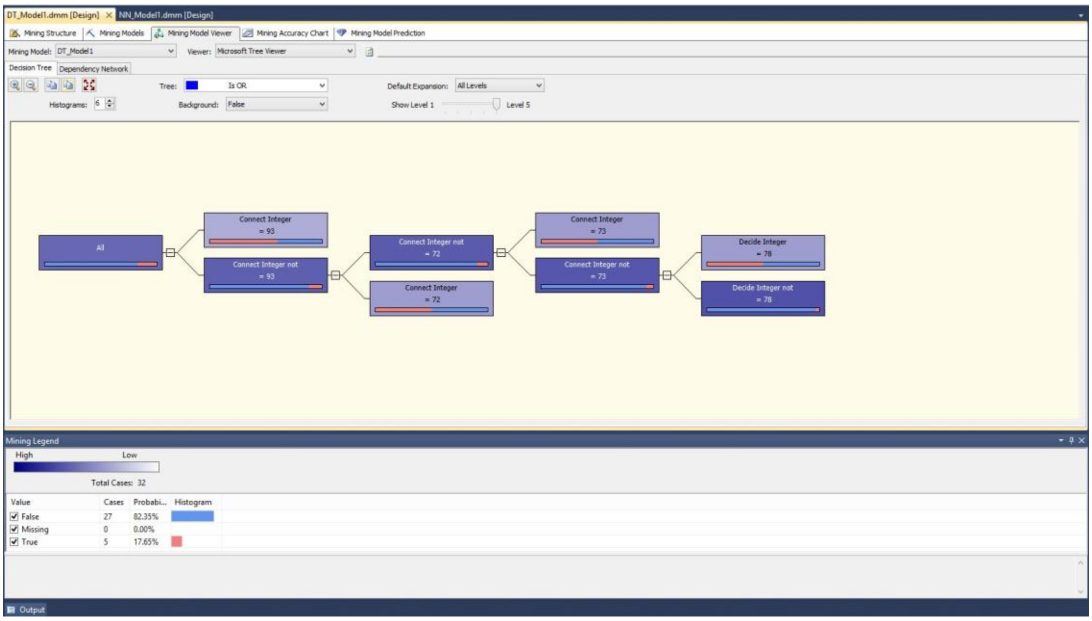
Show Level 1 Level 5

Mining Legend

Total Cases: 32

Value	Cases	Probabi...	Histogram
☑ False	27	82.35%	🟦
☑ Missing	0	0.00%	
☑ True	5	17.65%	🟥

Fig. A8.43: Example of the DT_Model1 built for ‘IsOR’ = True



Mining Model: DT_Model1

Viewer: Microsoft Tree Viewer

Tree: Is OR

Default Expansion: All Levels

Histograms: 6

Background: False

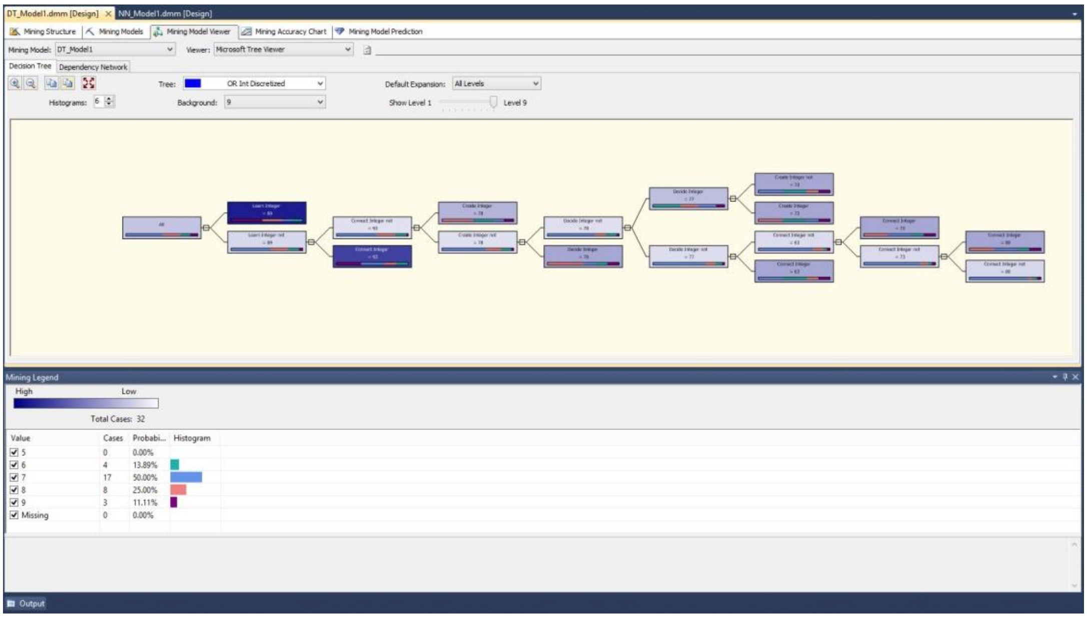
Show Level 1 Level 5

Mining Legend

Total Cases: 32

Value	Cases	Probabi...	Histogram
☑ False	27	82.35%	🟦
☑ Missing	0	0.00%	
☑ True	5	17.65%	🟥

Fig. A8.44: Example of the DT_Model1 built for ‘IsOR’ = False



Mining Model: DT_Model1 Viewer: Microsoft Tree Viewer

Tree: OR Int Discretized Default Expansion: All Levels

Histograms: 6 Background: 9 Show Level 1 Level 9

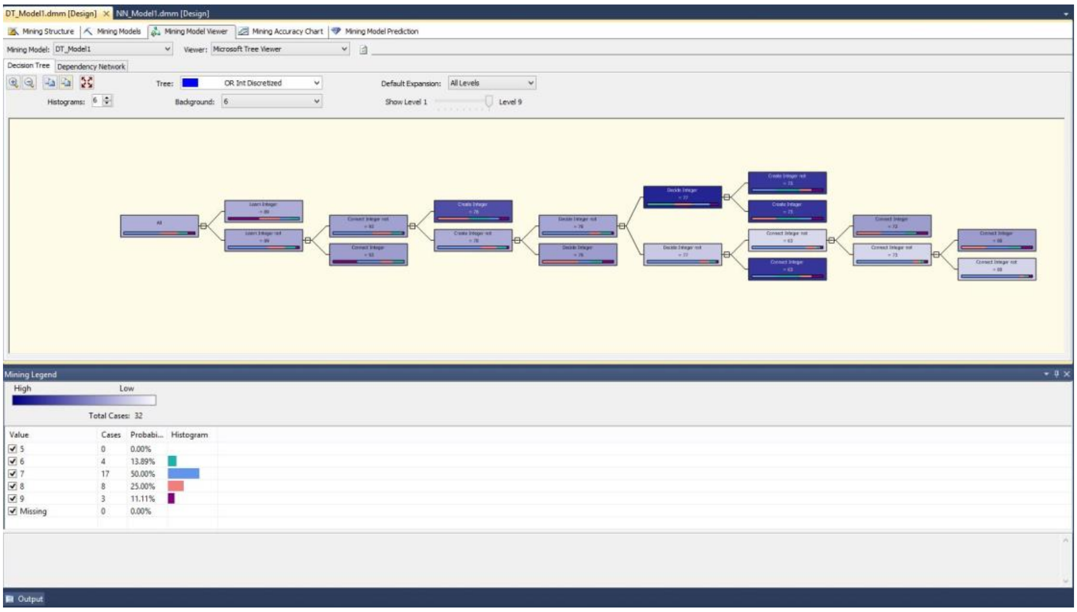
Mining Legend

High Low

Total Cases: 32

Value	Cases	Probabi...	Histogram
☑ 5	0	0.00%	
☑ 6	4	13.89%	▬
☑ 7	17	50.00%	▬
☑ 8	8	25.00%	▬
☑ 9	3	11.11%	▬
☑ Missing	0	0.00%	

Fig. A8.45: Example of DT_Model1 built for 'ORIntDiscretized' = 9 (the highest value)



Mining Model: DT_Model1 Viewer: Microsoft Tree Viewer

Tree: OR Int Discretized Default Expansion: All Levels

Histograms: 6 Background: 6 Show Level 1 Level 9

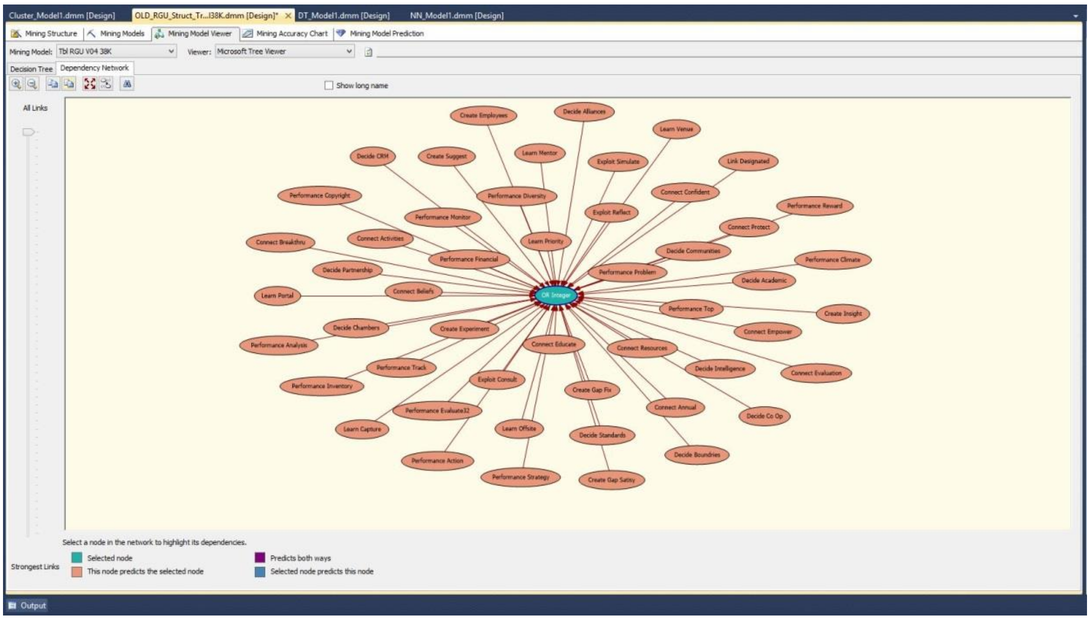
Mining Legend

High Low

Total Cases: 32

Value	Cases	Probabi...	Histogram
☑ 5	0	0.00%	
☑ 6	4	13.89%	▬
☑ 7	17	50.00%	▬
☑ 8	8	25.00%	▬
☑ 9	3	11.11%	▬
☑ Missing	0	0.00%	

Fig. A8.46: Example of DT_Model1 built for 'ORIntDiscretized' = 6 (the lowest value for which meaningful tree could be built)



Mining Model: Tbl RGU V04 38K

Viewer: Microsoft Tree Viewer

Select a node in the network to highlight its dependencies.

Strongest Links

🟦 Selected node

🟧 This node predicts the selected node

🟪 Predicts both ways

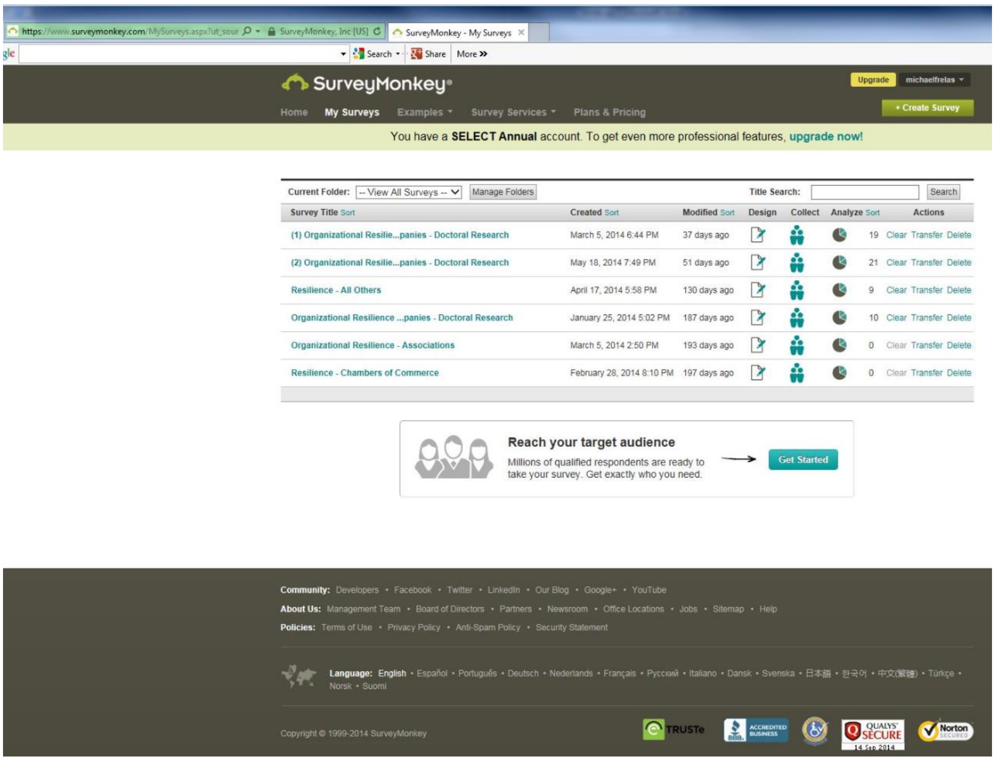
🟦 Selected node predicts this node

Fig. A8.47: Example of the network diagram when all KM activities are used as an input into the model

APPENDIX IX: Data Loading Steps

The following steps were carried out in order to load the responses to the questionnaire into the table in the MS SQL Server's database.

1. Files, in Excel format, were retrieved from survey Monkey on October 26, 2014. Here is pictorial representation of the files residing on the SurveyMonkey server.



You have a **SELECT Annual** account. To get even more professional features, upgrade now!

Current Folder: -- View All Surveys -- Manage Folders Title Search: Search

Survey Title Sort	Created Sort	Modified Sort	Design	Collect	Analyze Sort		Actions
(1) Organizational Resilie...panies - Doctoral Research	March 5, 2014 6:44 PM	37 days ago				19	Clear Transfer Delete
(2) Organizational Resilie...panies - Doctoral Research	May 18, 2014 7:49 PM	51 days ago				21	Clear Transfer Delete
Resilience - All Others	April 17, 2014 5:58 PM	130 days ago				9	Clear Transfer Delete
Organizational Resilience ...panies - Doctoral Research	January 25, 2014 5:02 PM	187 days ago				10	Clear Transfer Delete
Organizational Resilience - Associations	March 5, 2014 2:50 PM	193 days ago				0	Clear Transfer Delete
Resilience - Chambers of Commerce	February 28, 2014 8:10 PM	197 days ago				0	Clear Transfer Delete

Reach your target audience

Millions of qualified respondents are ready to take your survey. Get exactly who you need.

Get Started

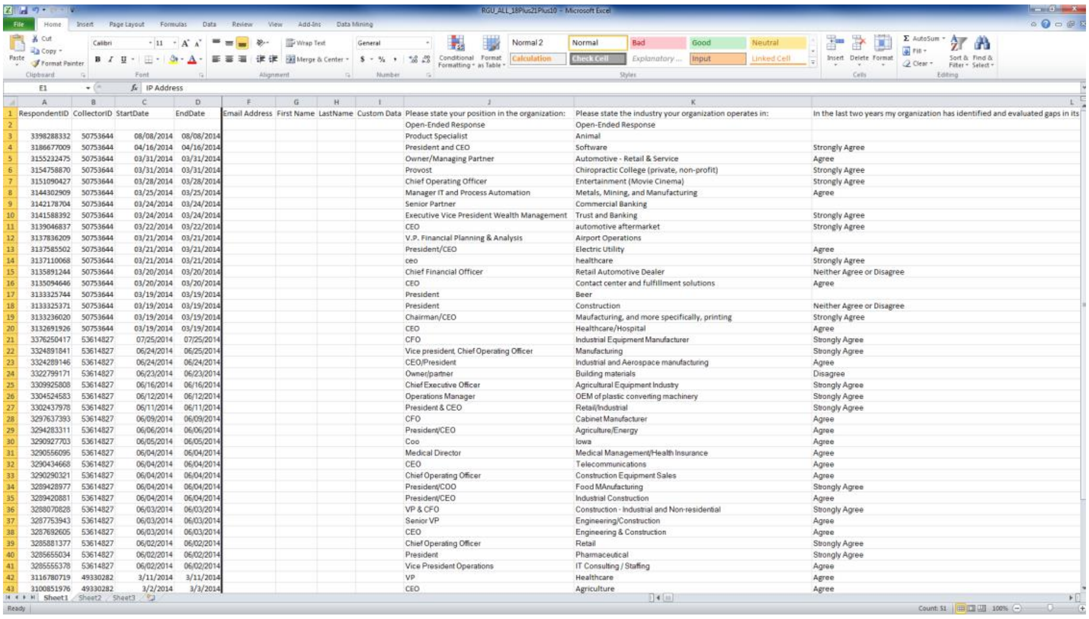
Community: Developers • Facebook • Twitter • LinkedIn • Our Blog • Google+ • YouTube
About Us: Management Team • Board of Directors • Partners • Newsroom • Office Locations • Jobs • Sitemap • Help
Policies: Terms of Use • Privacy Policy • Anti-Spam Policy • Security Statement

Language: English • Español • Português • Deutsch • Nederlands • Français • Русский • Italiano • Dansk • Svenska • 日本語 • 한국어 • 中文(繁體) • Türkçe • Norsk • Suomi

Copyright © 1999-2014 SurveyMonkey

Fig. A9.1: List of Excel files holding questionnaire replies.

All downloaded Excel files have been merged into a single Excel file. The Excel file has been formatted for further processing:



E1 IP Address

	A	B	C	D	F	G	H	I	J	K	L
1	RespondentID	CollectorID	StartDate	EndDate	Email Address	First Name	LastName	Custom Data	Please state your position in the organization:	Please state the industry your organization operates in:	In the last two years my organization has identified and evaluated gaps in its
2									Open-Ended Response	Open-Ended Response	
3	3398288332	50753644	08/08/2014	08/08/2014					Product Specialist	Animal	
4	3186677009	50753644	04/16/2014	04/16/2014					President and CEO	Software	Strongly Agree
5	3155232475	50753644	03/31/2014	03/31/2014					Owner/Managing Partner	Automotive - Retail & Service	Agree
6	3154758870	50753644	03/31/2014	03/31/2014					Provost	Chiropractic College (private, non-profit)	Strongly Agree
7	3151090427	50753644	03/28/2014	03/28/2014					Chief Operating Officer	Entertainment (Movie Cinema)	Strongly Agree
8	3144302909	50753644	03/25/2014	03/25/2014					Manager IT and Process Automation	Metals, Mining, and Manufacturing	Agree
9	3142178704	50753644	03/24/2014	03/24/2014					Senior Partner	Commercial Banking	
10	3141588392	50753644	03/24/2014	03/24/2014					Executive Vice President Wealth Management	Trust and Banking	Strongly Agree
11	3139046837	50753644	03/22/2014	03/22/2014					CEO	automotive aftermarket	Strongly Agree
12	3137836209	50753644	03/21/2014	03/21/2014					V.P. Financial Planning & Analysis	Airport Operations	
13	3137585502	50753644	03/21/2014	03/21/2014					President/CEO	Electric Utility	Agree
14	3137110068	50753644	03/21/2014	03/21/2014					ceo	healthcare	Strongly Agree
15	3135891244	50753644	03/20/2014	03/20/2014					Chief Financial Officer	Retail Automotive Dealer	Neither Agree or Disagree
16	3135094646	50753644	03/20/2014	03/20/2014					CEO	Contact center and fulfillment solutions	Agree
17	3133325744	50753644	03/19/2014	03/19/2014					President	Beer	
18	3133325371	50753644	03/19/2014	03/19/2014					President	Construction	Neither Agree or Disagree
19	3133236020	50753644	03/19/2014	03/19/2014					Chairman/CEO	Maufacturing, and more specifically, printing	Strongly Agree
20	3132691926	50753644	03/19/2014	03/19/2014					CEO	Healthcare/Hospital	Agree
21	3376250417	53614827	07/25/2014	07/25/2014					CFO	Industrial Equipment Manufacturer	Strongly Agree
22	3324891841	53614827	06/24/2014	06/25/2014					Vice president, Chief Operating Officer	Manufacturing	Strongly Agree
23	3324289146	53614827	06/24/2014	06/24/2014					CEO/President	Industrial and Aerospace manufacturing	Agree
24	3322799171	53614827	06/23/2014	06/23/2014					Owner/partner	Building materials	Disagree
25	3309925808	53614827	06/16/2014	06/16/2014					Chief Executive Officer	Agricultural Equipment Industry	Strongly Agree
26	3304524583	53614827	06/12/2014	06/12/2014					Operations Manager	OEM of plastic converting machinery	Strongly Agree
27	3302437978	53614827	06/11/2014	06/11/2014					President & CEO	Retail/Industrial	Strongly Agree
28	3297637393	53614827	06/09/2014	06/09/2014					CFO	Cabinet Manufacturer	Agree
29	3294283311	53614827	06/06/2014	06/06/2014					President/CEO	Agriculture/Energy	Agree
30	3290927703	53614827	06/05/2014	06/05/2014					Coo	Iowa	Agree
31	3290556095	53614827	06/04/2014	06/04/2014					Medical Director	Medical Management/Health Insurance	Agree
32	3290434668	53614827	06/04/2014	06/04/2014					CEO	Telecommunications	Agree
33	3290290321	53614827	06/04/2014	06/04/2014					Chief Operating Officer	Construction Equipment Sales	Agree
34	3289428977	53614827	06/04/2014	06/04/2014					President/COO	Food MAnufacturing	Strongly Agree
35	3289420881	53614827	06/04/2014	06/04/2014					President/CEO	Industrial Construction	Agree
36	3288070828	53614827	06/03/2014	06/03/2014					VP & CFO	Construction - Industrial and Non-residential	Strongly Agree
37	3287753943	53614827	06/03/2014	06/03/2014					Senior VP	Engineering/Construction	Agree
38	3287692605	53614827	06/03/2014	06/03/2014					CEO	Engineering & Construction	Agree
39	3285881377	53614827	06/02/2014	06/02/2014					Chief Operating Officer	Retail	Strongly Agree
40	3285655034	53614827	06/02/2014	06/02/2014					President	Pharmaceutical	Strongly Agree
41	3285555378	53614827	06/02/2014	06/02/2014					Vice President Operations	IT Consulting / Staffing	Agree
42	3116780719	49330282	3/11/2014	3/11/2014					VP	Healthcare	Agree
43	3100851976	49330282	3/2/2014	3/3/2014					CEO	Agriculture	Agree

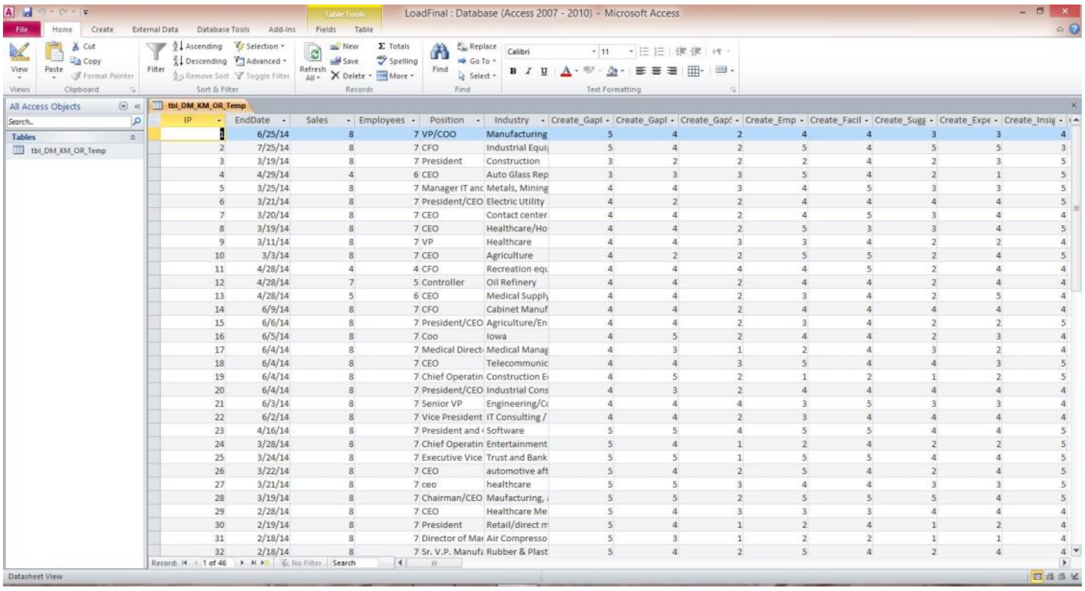
Fig A9.2: Excel file holding all responses to the questionnaire. (IP field has been hidden to protect privacy.)

- Column names have been changed; from the questionnaire questions to the shorter names.
- Responses have been changed, from categorical values ('Strongly Agree' and alike to 0 through 5).
- Additional columns have been inserted:
    - At the end of each section column holding the aggregated 'points' for specific question, per respondend. The name of such columns follows the following naming convention: 'Competence Area' and the word 'Points'. Example: 'ExploitPoints'.
    - The column holding maximum number achievable points for a given competence area. (The value in this column can possibly differ per respondent as any response 'N/A' decreases by 5 the value in this column for the respondent.) These column names will end with the 'PossiblePoints' string after the name of the competence area. Example: ExploitPossiblePoints.

  - The column whose name ends with 'Ratio', like 'ExploitRatio' holds the result of division of the number of points collected within a given competence are by the number of possible to achieve points. Example: ExploitRatio = ExploitPoints / ExploitPossiblePoints.
  - The 'Integer' column, like 'ExploitInteger' holds the value of the ratio field converted into the integer number type. Example: ExploitInteger will hold rounded up value of the ExploitRatio field.
  - The ending in 'Str' named column, like 'ExploitStr' will hold converted to categorical value the value of the 'Integer' field. That is, using the conversion rule described in Chapters 5 & 6, the 'ExploitStr' field will hold converted to categorical value the value of 'Integer' field. That is, 'ExploitStr' will hold the converted value of 'ExploitInteger' field.
  - The field 'Ratio7Areas' has been added to hold the values of total points achieved in seven categories (six competence areas plus performance section) divided by the total maximum possible number to achieve.
  - The field 'Integer7Areas' has been added to hold the total number of points achieved across seven categories.
  - The field 'Str7Areas' has been added but it has not been used.

2. Described in the point 3 above new fields have been populated with the data – per discussion in Chapters 5 & 6.

3. The contents of the Excel file have been loaded into Microsoft Access (2010) database, into the table: tbl_DM_KM_OR_Temp. (The MS Access-based table was used for quality check as it is easier to use table for such purpose than Excel file as the database queries can be created and executed against MA Access based database/table.)



tbl_DM_KM_OR_Temp

IP	EndDate	Sales	Employees	Position	Industry	Create_Gapl	Create_Gapl	Create_Gap!	Create_Emp	Create_Facil	Create_Sugg	Create_Expe	Create_Insig
1	6/25/14	8	7	VP/COO	Manufacturing	5	4	2	4	4	3	3	4
2	7/25/14	8	7	CFO	Industrial Equi	5	4	2	5	4	5	5	3
3	3/19/14	8	7	President	Construction	3	2	2	2	4	2	3	5
4	4/29/14	4	6	CEO	Auto Glass Rep	3	3	3	5	4	2	1	5
5	3/25/14	8	7	Manager IT anc	Metals, Mining	4	4	3	4	5	3	3	5
6	3/21/14	8	7	President/CEO	Electric Utility	4	2	2	4	4	4	4	5
7	3/20/14	8	7	CEO	Contact center	4	4	2	4	5	3	4	4
8	3/19/14	8	7	CEO	Healthcare/Ho	4	4	2	5	3	3	4	5
9	3/11/14	8	7	VP	Healthcare	4	4	3	3	4	2	2	4
10	3/3/14	8	7	CEO	Agriculture	4	2	2	5	5	2	4	5
11	4/28/14	4	4	CFO	Recreation equ	4	4	4	4	5	2	4	4
12	4/28/14	7	5	Controller	Oil Refinery	4	4	2	4	4	2	4	4
13	4/28/14	5	6	CEO	Medical Supply	4	4	2	3	4	2	5	4
14	6/9/14	8	7	CFO	Cabinet Manuf	4	4	2	4	4	4	4	4
15	6/6/14	8	7	President/CEO	Agriculture/En	4	4	2	3	4	2	2	5
16	6/5/14	8	7	Coo	Iowa	4	5	2	4	4	2	3	4
17	6/4/14	8	7	Medical Direct	Medical Manag	4	3	1	2	4	3	2	4
18	6/4/14	8	7	CEO	Telecommunic	4	4	3	5	4	4	3	5
19	6/4/14	8	7	Chief Operatin	Construction E	4	5	2	1	2	1	2	5
20	6/4/14	8	7	President/CEO	Industrial Cons	4	3	2	4	4	4	4	4
21	6/3/14	8	7	Senior VP	Engineering/Co	4	4	4	3	5	3	3	4
22	6/2/14	8	7	Vice President	IT Consulting /	4	4	2	3	4	4	4	4
23	4/16/14	8	7	President and	Software	5	5	4	5	5	4	4	5
24	3/28/14	8	7	Chief Operatin	Entertainment	5	4	1	2	4	2	2	5
25	3/24/14	8	7	Executive Vice	Trust and Bank	5	5	1	5	5	4	4	5
26	3/22/14	8	7	CEO	automotive aft	5	4	2	5	4	2	4	5
27	3/21/14	8	7	ceo	healthcare	5	5	3	4	4	3	3	5
28	3/19/14	8	7	Chairman/CEO	Maufacturing,	5	5	2	5	5	5	4	5
29	2/28/14	8	7	CEO	Healthcare Me	5	4	3	3	3	4	4	4
30	2/19/14	8	7	President	Retail/direct m	5	4	1	2	4	1	2	4
31	2/18/14	8	7	Director of Mai	Air Compresso	5	3	1	2	2	1	1	4
32	2/18/14	8	7	Sr. V.P. Manufa	Rubber & Plast	5	4	2	5	4	2	4	4

Fig. A9.3: Illustration showing the content of the intermediate table in MS Access.

4. As the last step of data loading the MS Access' table (tbl_DM_KM_OR_Temp) was copied, using the SSIS component named 'LoadTestData_RGU, into the tbl_DM_KM_OR table in SQL Server. The SSIS component simply copies data from the fields of the source table into the fields of the target table, performing no other function.

5. Data quality checks assuring the accuracy of the converted data were conducted by comparing the values in the terminal table with the data in the source Excel file (the file containing all questionnaire responses) and with the MS Access-based intermediate table. Because of the small number of records this process has been conducted manually and no problems were found.

Appendix X: Pilot Study Executive Letter

Dear Executive,

I recognize that the demands on your time are enormous, so my appreciation for your participation in this academic research project cannot be overstated. I am truly grateful for your time, and I hope to provide something of value for your organization in return for approximately 30 minutes of your time. This questionnaire is a chance for you to state your opinions for the benefit of mid-size businesses based in Midwest as well as the benefits of society.

My name is Michael Frelas. I am doctoral researcher studying the impact of knowledge management on organizational resilience within mid-size companies operating in the Midwest area of the US. [In short, I am trying to determine how successful companies are using and managing knowledge so that they stay at the top of their game.] While this work is conducted at a Scottish University (Robert Gordon University) I am a US citizen residing in the NW suburbs of Chicago.

By completing this questionnaire, you will be contributing to research in the field. Your input is of great value to this work and is greatly appreciated. In return for your time devoted to answering this questionnaire you will be provided (free of charge) with a feedback on your organization's performance vs. other participating companies. A free copy of my doctoral thesis will also be available. Please indicate if you wish to receive a copy at the end of this questionnaire. Please note, any responses you provide will be treated confidentially, and your anonymity will be preserved.

Finally, your reflection on the questionnaire's weak points (see the very last page) would be of extreme value to me and to this research. Please share your observations and/or opinions.

Once again, I would like to thank you for your time.

Please do not hesitate to contact me or my research supervisor if you have any questions or comments related to this research.

Best regards,

Michael Frelas, Doctoral Candidate

m.frelas@rgu.ac.uk

Cell phone #: (773) 505-8377

Research Supervisor:

Dr. Simon Burnett s.burnett@rgu.ac.uk

To make this questionnaire better please provide your feedback below.

Thank you very much again for your time!

6. Is this research questionnaire manageable in length?

   ____

7. Is this research questionnaire manageable in complexity?

   ____

8. Are the questions clear? If not, which questions need more clarification?

   ____

9. Do you feel that some of the questions were too general? If so, which questions appeared to be

   too general?

   ____

10. How can this questionnaire be improved?

   ____

   ____

   ____

   ____

   ____

   ____

   ____

   ____

   ____

APPENDIX XI: DM Structures

This appendix presents the data mining structures used by each data mining algorithm.

NB_Molde1 & NB_Model2:

DM structure name:	Source column name:	Comments:
Connect Str	ConnectStr	Categorical column, used as input/output for this model.
Create Str	CreateStr	Categorical column, used as input/output for this model.
Decide Str	DecideStr	Categorical column, used as input/output for this model.
Exploit Str	ExploitStr	Categorical column, used as input/output for this model.
IP	IP	Key column of integer type
Learn Str	LearnStr	Categorical column, used as input/output for this model.
Link Str	LinkStr	Categorical column, used as input/output for this model.
OR Int Discretized	ORIntDiscretized	Discrete, numerical column, used as input/output for this model.

Fig. A11.1: Columns from the tbl_DM_KM_OR_RGU table used by the Naïve Bayes algorithms.

Cluster_Model1:

DM structure name:	**Source column name:**	**Comments:**
Connect Activities	Connect_Activities	(Tiny) integer type of attribute, treated by the model as the continuous type of data.
Connect Beliefs	Connect_Beliefs	Same as above.
Connect Breakthru	Connect_Breakthru	Same as above.
Connect Buying	Connect_Buying	Same as above.
Connect Confident	Connect_Confident	Same as above.
Connect Educate	Connect_Educate	Same as above.
Connect Empower	Connect_Empower	Same as above.
Connect Evaluation	Connect_Evaluation	Same as above.
Connect Protect	Connect_Protect	Same as above.
Connect Relations	Connect_Relations	Same as above.
Connect Resources	Connect_Resources	Same as above.
Create Employees	Create_Employees	Same as above.
Create Experiment	Create_Experiment	Same as above.
Create Facilities	Create_Facilities	Same as above.
Create Gap Fix	Create_Gap_Fix	Same as above.
Create Gap Id	Create_Gap_Id	Same as above.
Create Gap Salary	Create_Gap_Salary	Same as above.
Create Insight	Create_Insight	Same as above.
Create Suggest	Create_Suggest	Same as above.
Decide Academic	Decide_Academic	Same as above.
Decide Alliances	Decide_Alliances	Same as above.
Decide Boundries	Decide_Boundries	Same as above.
Decide Chambers	Decide_Chambers	Same as above.
Decide Co Op	Decide_Co_Op	Same as above.
Decide Communities	Decide_Communities	Same as above.
Decide Condition	Decide_Condition	Same as above.
Decide CRM	Decide_CRM	Same as above.
Decide Intelligence	Decide_Intelligence	Same as above.

Decide Partnership	Decide_Partnership	Same as above.
Decide Professional	Decide_Professional	Same as above.
Decide Standards	Decide_Standards	Same as above.
Exploit Consult	Exploit_Consult	Same as above.
Exploit Electronic DB	Exploit_Electronic_DB	Same as above.
Exploit References	Exploit_References	Same as above.
Exploit Reflect	Exploit_Reflect	Same as above.
Exploit Simulate	Exploit_Simulate	Same as above.
Learn BI	Learn_BI	Same as above.
Learn Capture	Learn_Capture	Same as above.
Learn Mentor	Learn_Mentor	Same as above.
Learn Offsite	Learn_Offsite	Same as above.
Learn Portal	Learn_Portal	Same as above.
Learn Priority	Learn_Priority	Same as above.
Learn Reimburse	Learn_Reimburse	Same as above.
Learn Training	Learn_Training	Same as above.
Learn Venue	Learn_Venue	Same as above.
Link Actively	Link_Actively	Same as above.
Link Designated	Link_Designated	Same as above.
Link Leadership	Link_Leadership	Same as above.
Link Monitor	Link_Monitor	Same as above.
Link Outsourcing	Link_Outsourcing	Same as above.
Link Relationship	Link_Relationship	Same as above.
OR Integer	OR_Integer	Same as above.
IP	IP	Key column of integer type

Fig. A11.2: Based on the tbl_DM_KM_OR_RGU table clustering DM structure.

NN_Model1:

DM structure name:	**Source column name:**	**Comments:**
Connect Integer	ConnectInteger	Continuous type of variable used for input

		in this model.
Create Integer	CreateInteger	Continuous type of variable used for input in this model.
Decide Integer	DecideInteger	Continuous type of variable used for input in this model.
Exploit Integer	ExploitInteger	Continuous type of variable used for input in this model.
IP	IP	Key column of integer type
Learn Integer	LearnInteger	Continuous type of variable used for input in this model.
Link Integer	LinkInteger	Continuous type of variable used for input in this model.
OR Integer	ORInteger	Continuous type of variable used for output in this model.

Fig. A11.3: Based on tbl_DM_KM_OR_RGU table definition of the NN_Model's DM structure.

DT_Model1:

DM structure name:	Source column name:	Comments:
Connect Integer	Connect_Integer	Column was designated as of discrete, instead of continuous type.
Create Integer	Create_Integer	Same as above.
Decide Integer	Decide_Integer	Same as above.
Exploit Integer	Exploit_Integer	Same as above.
Learn Integer	Learn_Integer	Same as above.
Link Integer	Link_Integer	Same as above.
IP	IP	Key column of integer type.
Is OR	IsOR	Discrete, Boolean (Yes/No) type.
OR Int Discretized	ORIntDiscretized	Discrete integer.

Fig. A11.4: DT model structure.

APPENDIX XII: DM Parameters

This appendix presents the data mining parameters used by each data mining algorithm.

Cluster_Model1:

Parameter:	**Description:**	**Use in this research:**
CLUSTERING_ METHOD	Indicates which algorithm is used to determine cluster membership. 1 = Scalable EM; 2 = Non-scalable EM; 3 = Scalable K-Means; 4 = Non-scalable K-Means.	Per discussion in Chapter 6.3, scalable algorithms will not be used as those are primarily designed to be used with large data sets. This research uses the value of 2 and 4 for this parameter only.
CLUSTER_ COUNT	Specifies the approximate number of clusters to find.	This research uses two values for this parameter. The value of 0, allowing the algorithm to choose the number of segments and 6, to correspond to the count of six competencies.
CLUSTER_ SEED	The random number that is used to initialize the clusters.	The default value of 0 has not been changed. (For testing purposes

		this value can be changed to make sure resultant models do not vary greatly indicating model’s stability.)
MINIMUM_ SUPPORT	Specifies the number of cases that are needed to build a cluster.	With the very limited amount of data the value of this variable is left set at a default value of 1.
MODELLING_ CARDINALITY	This parameter controls how many candidate models are generated during clustering.	The reduction of the number of this parameter can carry the potential cost of accuracy. However, with the limited amount of data available, the default value of 10 has been used as the value for this parameter.
STOPPING_ TOLERANCE	This parameter is used to determine when the model has converged and the algorithm has finished building the model. (It represents the maximum number of cases that can change membership before the model is considered as converged.)	With only approximately 32 input data elements and potential six segments, the value of this parameter has been changed from the default value of 10 to 4.
SAMPLE_SIZE	Specifies the number of cases	The value of this

	used in each step (affects only scalable clustering methods). Can be used for 'quick clustering' for a large data set but at a risk that not all of the cases will be considered in modelling.	parameter has no effect on the model used in this research as non-scalable methods are used in the research.
MAXIMUM_INPUT_ ATTRIBUTES	Specifies the maximum number of attributes that the algorithm can handle before automatic feature selection (selecting the 'most popular attributes') is invoked. (The higher the number, the lower performance.)	With the default value set to 255, the number addresses all the needs of this research.
MAXIMUM_STATES	Controls how many states one particular attribute can have.	The default value of 100, even though highly excessive, was left unchanged.

Fig. A12.1: Parameters for Cluster_Model1 model.

NN_Model1:

Parameter:	Description:	Use in this research:
MAXIMUM_INPUT_ ATTRIBUTES	Specifying the maximum number of input attributes that the algorithm can handle before invoked implicitly to pick the most significant attributes.	Used default value of 255.
MAXIMU_OUTPUT_ ATTRIBUTES	Similar to the above but for output attributes.	Used default value of 255.
MAXIMUM_STATES	Specifies the maximum number of attribute states that the algorithm supports.	Used default value of 100.
HOLDOUT_ PERCENTAGE	Specifies the percentage of cases within the training data used to calculate the holdout error for the algorithm.	Used default value of 30.
HOLDOUT_SEED	An integer that specifies the seed for selecting the holdout data set.	Used default value of 0. (Algorithm generated seed for the holdout.)
HIDDEN_NODE_ RATIO	Specifies a number used in determining the number of nodes in the hidden layer.	Used default value of 4.
SAMPLE_SIZE	Is the upper limit of the number of cases used for training.	Used default value of 10000.

Fig. A12.2: Parameters for NN_Model1 model.

DT_Model1:

Parameter:	Description:	Use in this research:
COMPLEXITY_PENALTY	Inhibits the growth of the decision tree. Decreasing the value increases the likelihood of a split while increasing the value decreases the likelihood.	Per specifications in the modelling software: for number of parameters between 1 and 9 the value was set to 0.5.
FORCE_REGRESSOR	Forces the algorithm to use the indicated columns as regressors in the regression formula regardless of their importance as calculated by the algorithm.	N/A as this work does not use regression algorithm.
MAXIMUM_INPUT_ ATTRIBUTES	Specifies the maximum number of input attributes that the algorithm can handle before invoking feature selection (selecting the most important attributes).	Used default value of 255.
MAXIMUM_OUTPUT_ ATTRIBUTES	Same as above but for output attributes.	Used default value of 255.
MINIMUM_SUPPORT	Specifies the minimum number of cases that a leaf node must contain.	Set value to 2 as the value of 1 specifies the minimum number of cases as a percentage of the total cases.

SCORE_METHOD	Specifies the method to calculate the split score. The available methods include: Entropy(1), Bayesian with K2 Prior (3), or Bayesian Dirichlet Equivalent with Uniform prior (4)	Needed to use the value of 1 in order to obtain a binary tree. Other methods produced less desired results – described further in this section.
SPLIT_METHOD	Specifies the method used to split the node. The possible choices are: Binary (1), Complete (2), or both (3).	Value of 1 was chosen as all other values (regardless of other parameters values) would produce no tree.

Fig. A12.3: Parameters for DT_Model1 model.