SOUNDEARAJAH, S. and ASANKA, P.P.G.D. 2022. Sentiment analysis of ASOS product reviews using machine learning algorithms by comparing several models. In *Proceedings of 2022 International research conference on Smart computing and systems engineering (SCSE 2022), 1 September 2022, Colombo, Sri Lanka*. Hosted on IEEE [online], pages 143-150. Available from: https://doi.org/10.1109/scse56529.2022.9905147

Sentiment analysis of ASOS product reviews using machine learning algorithms by comparing several models.

SOUNDEARAJAH, S. and ASANKA, P.P.G.D.

2022

Sentiment Analysis of ASOS Product Reviews Using Machine Learning Algorithms by Comparing Several Models

Sahithya Soundearajah*
School of Computing
Robert Gordon University Scotland, United Kingdom
s.soundearajah@rgu.ac.uk

P. P. G. Dinesh Asanka
Department of Industrial Management
Faculty of Science, University of Kelaniya, Sri Lanka
dasanka@kln.ac.lk

***Abstract* - Digital ratings are crucial in improving international customer communications and impacting consumer purchasing trends. To obtain important data from a massive number of customer reviews, they must be sorted into positive and negative opinions. Sentiment analysis is a computational method for extracting emotive information from a text. In this particular research, over 3000 reviews have been obtained from the ASOS website and classified into three different sentiments: excellent, average, and bad. The obtained reviews have been pre-processed, then feature extraction is applied to the pre-processed data to remove the redundant data. Finally, distinct machine learning algorithms will be utilized to build disparate models. This research is vital as it allows the ASOS organization to gain insight into how consumers perceive about specific issues and detect urgent issues such as delivery delays and misplaced packages in the current time period before the issue goes outof control. The key results of this research show that the Nu- Support Vector Classification model obtained the highest accuracy score of 85.99% and the lowest accuracy score of 51.47% was obtained for the AdaBoost classifier model.**

Keywords – feature extraction, Machine Learning algorithms, multi-class classification, sentiment analysis

I. INTRODUCTION

With the fast expansion of e-commerce, online purchasing has increased. Due to the information inconsistency between the real product quality and the description provided by the seller, more and more consumers seek product information from electronic commerce reviews that involve numerous features such as prices, services and operations [1]. Sentiment analysis is frequently used in business to examine and predict the view or conduct of a specific group.

Current brands constantly monitor social media and other online distribution platforms [2]. Each source of data will offer distinct viewpoints on the products and label, providing the knowledge needed to make smarter e-commerce choices. Global customers who purchase items on the ASOS website have the opportunity to write about and share both positive and bad experiences [3]. It is expected that the organization should accept and appreciate user sentiments; nevertheless, this poses a problem when a review is nasty or breaches community standards. However, if users praise the delivery companies and products associated with ASOS, it will cause a rise in the number of customers who shop on the ASOS website. In this research, the final objective is to identify the correct parameter to assess the usability of sentiment analysis.

The following research questions were chosen, keeping in mind that the quick growth of online systems has intensified in the past few years due to the focus of product reviews on sentiment analysis technology.

A. Research Questions and Objectives

The following research questions have been stated for this research. The first research question is to review the customer feedback available on the ASOS website using a sentiment analysis method by comparing several different models. The second research question is to use the different models created to predict the sentiment of random reviews on the ASOS website by utilizing the multi-class classification method and the final research question is to find any relationship between the multi-class classification method and the keyword. The research aims to classify the reviews on the basis of the sentiment of the customers, hence providing sentiment orientation of the review, resulting in better judgment.

The first objective of this research is to analyze each customer review and place each review in the appropriate class, thereby enabling the ASOS website to improve service quality and customer satisfaction. The second objective is to analyze how keywords can be obtained from the dataset after feature extraction. The third objective is to identify the best sentiment analysis model from a range of different models, after applying different machine learning algorithms. The final objective is to prove that all the models created can predict the correct class of any random review found on the ASOS website.

Multi-class sentiment analysis enables the individual who wishes to purchase a product on the online website to obtain an honest opinion about the products from several people of different cultures, ages, nationalities and occupations. Predictive data analytics based on machine learning can predict the sentiment of new reviews posted by each customer on the ASOS website, to identify whether the review has been classified into the correct class [4].

The customer reviews collected are the thoughts of real consumers, so all the noise is removed. It provides a clear picture of the company's operations, as the customer states their opinion based on the type of product bought, the condition that they received the clothes and the time taken for the delivery [5]. The results of sentiment analysis on reviews can indicate particular product performance, reveal gaps in predictions, and deliver other significant research information [6].In summary, applying sentiment analysis on product evaluations offers more insights about product performance.

and it reveals how the products sold on the ASOS website and the clothes distribution influence the ASOS brand perception. Through sentiment analysis, the product issues can be detected at an early stage to identify whether the clothes sold by ASOS are adored and purchased by a significant number of online users. Prevention is better than cure, and it helps the company investigate if there are more items being returned or more items being returned or more purchases being made each month [2].Hence, predicting the class of new customer reviews posted on the ASOS website through the different models created enables the company to address the problems written by the customers and enhances customer satisfaction.

To conclude, analyzing online reviews using techniques like sentiment analysis enables to boost the organization's sales and adherence to the issues stated by the online customers [3]. Moreover, sentiment analysis involves selecting the text classification and applying several pre-processing methods, to determine the intention of the author of the text, thereby understanding whether the online purchaser intends to criticize or admire the items sold on the website [7]. Through multimodal sentiment analysis, the product features of items can be improved, and the quality of the merchandise can be advanced, by comparing the competitor product reviews and providing a better-quality end product for the consumers. A brief review of some related work on sentiment analysis has been provided in the literature review section. The data collection methods and different pre-processing techniques have been explained in the methodology portion. The different models being built have been described in the design of the solution section and the method of applying the models has been explicitly discussed in the implementation section. The conclusion section contains a brief description of the contribution of the research as well as the future work that can be done to improve this research.

II. LITERATURE REVIEW

A. *Theories and Critical Review*

This paper will examine different machine learning algorithms like support vector machine, naïve Bayes algorithms and K-nearest neighbor algorithm applied on raw data like customer feedback and Twitter data to obtain critical information for sentiment analysis. The research work included below supports the research topic and describes how different machine learning algorithms can be applied to the text after pre-processing the text to obtain an optimum model.

Two generative models, MaxEnt-JABST and JABST, were utilized to extract fine-grained thoughts as well as elements of customer feedback. The JABST model collected specific and generalized views along with features of emotion orientation [8]. Moreover, the MaxEnt-JABST design also included a maximal entropy classifier for accurately differentiating features or different viewpoints. Furthermore, two machine learning techniques, namely naïve Bayes classification and support vector machine techniques, were used to conduct sentiment analysis on certain customer reviews based on the products bought [9]. An Amazon dataset was utilized to extract opinion lexicons, which included 4783 negative and 2006 positive terms with sentiment ratings for each phrase. A text mining technique was utilized on a vast dataset of product evaluations to analyze customer reviews, identify consumer opinions, and perform sentiment analysis submitted by disparate customers on the internet [10]. Key graph keyword extraction was used to extract keywords from each document with high-frequency provided by several online users [10]. Key graph keyword extraction was employed to collect keywords from each article including high-frequency phrases and to evaluate the strength of sentiment polarity. To classify data based on sentiment strength, k-means clustering was used. An aspect-level sentiment analysis method was identified by identification, aggregation, and classification [11]. Parts-of-speech tagging was used to extract adjectives from the sentences.

A Research based on emotion classification algorithms, support vector machine algorithms and latent sentiment analysis methods was carried out in 2018 [12]. The latent sentiment analysis was used to optimize the extraction of features for text, the support vector machine algorithm was employed as a classifier to improve emotional classification precision and computational productivity. The results demonstrated that the model could effectively estimate online feedback. A crossbred attribute-centric sentiment categorization was introduced to incorporate domain-specific knowledge and extract definite word relations [13]. Datasets were created to test the effectiveness of the method.

A ranking technique based on website reviews on the many aspects of different products, including the personal feelings of customer ratings, was investigated [14]. The sentiment value of the product was principally assessed by determining the ratings of specific aspects using the Latent Dirichlet allocation approach. The total result of each product was evaluated using the page rank method, and a graph was produced. A research [15] has suggested a novel method for encoding sentiment analysis findings using interval type-2 fuzzy integers that took accuracy rates into account. The sentiment analysis findings with a 100% accuracy rate were turned into a triangular fuzzy number, while those with lower accuracy rates were treated as the membership function's uncertainty. It was proposed to use a multi-model combined sentiment theme model [16]. Using additional previous knowledge samples, support vector machines were employed to categorize stock forum postings [17]. Another research [18] has suggested the importance of machine learning in the e-commerce area and the use of cloud platforms for predictive analysis. The decision table was one of the top classifiers with high accuracy among the eleven data mining classification approaches studied [1].

B. *Gap Analysis*

The different experimental outcomes and unique contributions of each research are identified in the gap analysis. Knowledge obtained from the gap analysis has been applied in the research methods and research design section.

Another research [8] has asserted that the experimental outcomes demonstrated that the designs surpassed current benchmarks and could comprehend fine-grained characteristics and views, but further improvement was required. A research [9] has affirmed that the support vector machine renders lower accuracy in classification. Moreover, emotion recognition is an extraction process that helps to recognize the mood and perception of the individual; however, emotion mining techniques were not used to classify the reviews and understand the customers' mindset. Another research [10] has employed the sentiment analysis technique that assessed sentiment polarity at the document level, instead of evaluating sentiment polarity at the phrase level. As k-means clustering was used, it may bring about over clustering. Using aspect-level sentiment analysis, the classification was executed by utilizing two machine algorithms, naïve Bayes, and support vector machine classification [11]. The outcomes revealed that it attained more accurateness from the naïve Bayes algorithms when compared with the support vector machine classification algorithms. The support vector machine classification algorithm approach was not apt for large datasets.

The emotion classification approach affirmed by [12] is not efficient, as it uses only a lesser amount of data for analysis. The hybridized attribute-centric sentiment classification stated by [13] discovered the most prevalent bigrams and trigrams in the corpus. The method has proved unsuccessful in detecting the attribution and demanded a prolonged computational effort. Although sentiment classification was carried out, emotion mining was not applied to identify the opinion words in the dataset. [14] stated that the system used the Latent Dirichlet allocation method, which was sensitive to overfitting, and validation of Latent Dirichlet allocation models were problematic. The system's performance was estimated through the appropriate theoretical investigation offered for the constructed interval type-2 fuzzy numbers, and it rendered a better outcome [15]. To analyze concealed user emotions and theme classifications in Weibo text, the model built by [16] has utilized a Latent Dirichlet allocation method. Although the machine-learning-based technique may select the characteristics automatically, it relies largely on manual feature selection. [17]'s deep-learning-based approach requires no manual involvement at all. It can choose and retrieve attributes automatically using the neural network structure and benefit from its own failures. [18] proposed that two predictive paradigms were built on real cloud platforms using popular machine learning classification algorithms: multi-class decision forest and multi-class logistic regression. The performance of both the models were evaluated on basis of classification accuracy on one percent data. The dataset chosen by [9] is composed of an ordering log file for three months. The paper states the trial results that were conducted with different feature classification methods in combination with three main classifiers: naïve Bayes, support vector machine, K-nearest neighbors, along with LDA, an unsupervised document topic classifier.

Lastly, using sentiment analysis, the gap analysis helps to understand the new predictive analytic methods researched to obtain more meaningful information from customer feedback, ratings and Twitter data. The gap analysis assists the research by identifying unique key methods that can be applied to the ASOS reviews by applying different pre-processing techniques that help remove redundancies and eliminate the inconsistencies and duplicates in the data obtained. Methodology' states the specific methods of data collection and encompasses the different approaches in which the research is carried out by applying feature extraction to obtain the important keywords and applying different machine learning algorithms to identify the best model.

III. METHODOLOGY

The main research problem identified in this research is how certain companies can improve their services when users purchase items on the company website by analyzing the reviews written by customers after acquiring the products and stating their opinion based on the customer experience, delivery time and the quality of how the product is delivered to the customer.

A. *Research Design*

A qualitative design was chosen for this research because qualitative research involves gathering and analyzing non-numerical data to understand opinions and experiences [11]. In this research, qualitative data identifies the customers' personal feelings, through the comments provided and helps the company truly understand the customers' needs. Secondary data was collected for this research, as three thousand customer reviews were chosen from the ASOS website based on the customer rating provided by each customer for the three classes: excellent, bad and average. To ensure the reliability of the data, the customer feedback was only chosen from the official ASOS website and only English reviews were chosen. To maintain validity, a thousand reviews were chosen from each class and all the data selected were between January 2021 to August 2021.

B. *Data Collection*

A three-class classification had been applied to the dataset based on the customer reviews classification found on the ASOS website. The reviews had been separated into three classes: excellent, bad and average. The data was entered into a Microsoft Excel spreadsheet and saved in the format of a comma-separated values file. Only the English reviews were selected from the ASOS website, however, reviews from different countries were selected and for each class, 1000 reviews were obtained.

C. *Pre-Processing*

The purpose of pre-processing methods is to convert the data into a more manageable format for the feature extractor and also to remove superfluous information. The following pre-processing data techniques will be carried out using Python program for this research, such as lemmatization, stemming and removal of stop-words, URLs, tokenization, special characters, and numbers. Firstly, the pre-processing step carried out is tokenization, essentially dividing the text into a set of significant chunks known as tokens. Stop-words like 'is' are removed so that the focus is given to only the important information in the reviews, and it improves the model's performance, without changing the semantics of the text. Lemmatization decreases the number of distinct occurrences of identical text tokens by transforming multiple relevant keywords intoa single canonical form [15]. For instance, the words 'changes', 'changing', and 'changed' will all be transformed to 'change.' The procedure of stemming is employed to obtain the basic form.

D. *Feature Extraction*

As the pre-processed text obtained is both highly dimensional and unstructured, one can first extract the most distinguishing features of the text, thus reducing dimensionality. This process is called feature extraction. The first feature extraction method applied to the data is Bag of Words. The term Bag of Words means that models do not consider word order [5]. Another approach is to have a binary vector instead, keeping track of whether or not a word exists within a document. However, this approach also loses the multiplicity of the words in addition to the order and grammar. Next, the parts-of-speech tags feature was carried out to derive important nouns and adjectives from the pre-processed text as these words are good indicators of subjectivity and sentiment [8]. Thirdly, opinion words and phrases were selected as a phrase has a positive semantic tendency when it addresses a pleasant comment, for instance, “best customer service” and a negative semantic predisposition when it addresses a bad review, for instance, “poor customer service.”

Lastly, Term Frequency-Inverse Document Frequency (TF- IDF) is implemented on the dataset. TF-IDF is a strategy for extracting data that may be used to identify the significance of words in documents in reference to a query. [18]. In this case, it can be used for feature extraction by determining which terms in a document are most distinguishing for that document. TF- IDF consists of two steps, first calculating the term frequency (TF) and then calculating the inverse document frequency (IDF). TF can be calculated as how many times a term occurs in a document, just as how a vector is counted [2]. IDF is calculated by taking the total number of documents in the corpus and dividing it by the number of documents where the term appears. The result is then logarithmized. By multiplying the TF part and the IDF part for a certain term,a measure of how distinguishing that term is can be obtained. In this particular research, from the corpus of pre-processed reviews, the word “service” would probably have a pretty high TF-IDF score since it would often occur in the documents related to the processed reviews. In contrast, a common word like “able” has a low TF-IDF score of 0.0, since the word “able” is not specifically related to the delivery service offered by the ASOS company or the customer service provided on the ASOS website.

E. Machine Learning Algorithms

The following machine learning algorithms will be applied in this research to create different models from the pre-processed dataset. The best model can be identified based on the highest accuracy score obtained as well as by comparing the f1, recall and precision values. Additionally, to ensure that all the models are reliable, the dataset is split into 80: 20 for the train and test data.

A few examples of machine algorithms used to create the models were Random Forest algorithm; which is a classification and regression method based on the collection of a proliferation of decision trees [15]. Support vector machines are considered universal learners and were initially conceived for binary classification, but research has developed it into a multi-class classification [17], and the decision tree classifier which has been extensively used for predictions and classification.

F. Conceptual Framework

The conceptual framework in this research is used to show the relationships between the independent and dependent variables, how they relate to each other and the research study. For this specific research, the conceptual model shown below in Figure 1 demonstrates that the independent variables are text and country. The review text is customer feedback obtained from the ASOS website and the keywords are further extracted from the review text using the feature extraction method. Country is the specific location from where the review was posted on the ASOS website. As the research is on sentiment analysis, each of the customer reviews is separated into three distinct customer ratings which are bad, average and excellent. The three different categories are the dependent variable.

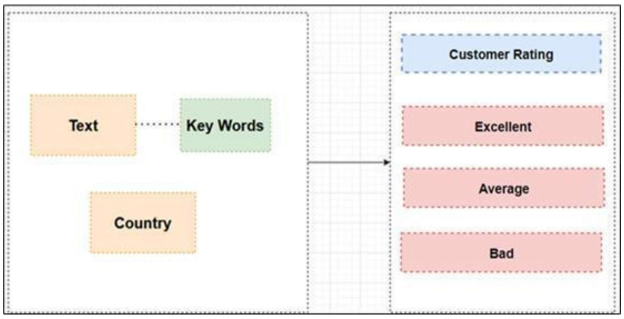

Fig. 1. Conceptual Model for 3 class classification models

In conclusion, customers' dependency on online recommendations has grown substantially due to an evolutionary transition from offline to electronic markets. As a result, the purpose of this study is to accomplish this by conducting sentiment analysis on clothing feedback and classifying the reviews into three distinct categories, with an equal ratio of thousand reviews for each sentiment and applying different machine learning algorithms such as naïve Bayes, support vector machine and decision tree, to create classification models to categorize the reviews.

IV. INTERPRETATION OF DATA

The procedure of ascribing interpretation to the information acquired after pre-processing and feature extraction, in addition to establishing the implications and consequences of the results, is known as data analysis and interpretation [4]. This section will discuss the visual representations obtained after applying different natural language processing techniques to show their importance for this research.

Important words were first extracted after applying different pre-processing techniques such as stemming as well as by extracting the nouns and adjectives after feature extraction. The sklearn.feature extraction library in Python can be utilized to derive characteristics from datasets comprising of customer feedback in a structure enabled by machine learning techniques [2]. Both bigram and trigram feature extraction were carried out. For example, “best customer service” is the most important keyword in the datasets obtained after trigram feature extraction as it had the highest frequency in the positive class: excellent. The vital keyword “worst customer service” had obtained the highest negative frequency according to the negative class: bad. Hence, trigram feature extraction is better than bigram feature extraction as it provides more resourceful keywords, unlike the bigram feature extraction which just provides keywords like “live chat” and “touch get”, which have no connection to the sentiment of the customer and does not enable to understand the emotion and sentiment of the consumer, unlike the trigram feature extraction.

A word cloud was created after pre-processing using Python software in Google Collaboratory. The packages “wordcloud” and “matplotlib.pyplot” were used to create Figure 2. The keyword ‘service’ is the largest word in the word cloud, as it has appeared 1408 times in the dataset. From the reviews obtained from the ASOS website, all the customers spoke about the different kinds of services received from ASOS.

Fig. 2. Word cloud created after pre-processing and feature extraction

V. DESIGN OF THE SOLUTION.

In the "Design of the Solution" chapter, the researcher states the different machine algorithms that have been applied to the pre-processed dataset. It also provides a classification for selecting each machine algorithm and highlights the key facts of the machine algorithm.

A. Deliverables

The dataset was created by obtaining customer reviews from the ASOS website and each review was classified into three classes. The data is pre-processed through various methods such as stemming, tokenization and lemmatization to remove any null values and inconsistencies. Feature extraction is utilized to derive characteristics from datasets comprising of customer feedback in a structure enabled by machine learning techniques. Next, the dataset is divided into two subgroups: the training and testing set [7]. Generally, a fraction of the dataset is taken to train the model. Here the datasets are divided into the ratio of 20:80 (i.e., test: train). To build the models, several machine learning techniques are employed and tested iteratively in the next step, and the best model is determined. The best model is chosen to generate more useful results based on the accuracy score of each model as well as recall, f1 values and precision values obtained from the classification report of each model.

B. Software Tools Used

The eighteen models were built using the Google Collaboratory application using Python. Different libraries such as "seaborn", "pandas", "matplotlib", "nltk", "WordCloud" and "sklearn.metrics" will be installed into the Python program for tokenization, text stemming, to generate word clouds, implementing score and utility functions to measure classification performance [10]. Various machine learning algorithms such as decision forest, support vector machine and XGBoost will be applied to create eighteen models [10]. The results obtained ensured that Nu-Support Vector Classification model provided the highest accuracy score of 85.99%.

The first model created for the three-class classification dataset is the Nu-Support Vector classification model. The nu-support vector classifier is akin to the Support Vector Machine algorithm with the only difference being that the nu-Support Vector Classifier has a new variable to administer the number of support vectors [13]. The second model created is the Linear Support Vector classification model. The Linear Support Vector Classification model is utilized for linearly separated data, which implies that if a dataset can be categorized into two categories. Whenever data is divided into classes by applying one straight line, it is referred to as linearly separable data, and the classifier employed is alluded to as a Linear Support Vector Classification classifier [2].

The third model created is the Multinomial Naïve Bayes Model. The sklearn.naïve_bayes.MultinomialNB classifier is appropriate for categorization using different features, such as a number of words in text classification [14]. The multinomial distribution represents the likelihood of witnessing counts in a variety of categories, multi nomial naïve Bayes is best suited for characteristics that indicate counts or count rates. A Decision Tree Classifier was applied to create the fourth model. Decision Tree is a supervised machine learning system that makes decisions based on a set of rules [17]. According to [18], the logic underlying Decision Trees is that the dataset characteristics are utilized to generate Boolean features: yes or no questions, and the dataset is repeatedly separated until all of the datapoints belonging to each class have been isolated. The data is structured into a tree structure as a result of this procedure.

The fifth model was generated using an AdaBoost Classifier. AdaBoost is an approach for recurrent ensemble construction. Adaboost classifier's core principle is to establish the strengths of classifiers and instruct the data sample in each repetition to guarantee correct estimates of unexpected occurrences [5]. The sixth model was created using the Bernoulli naïve Bayes classifier. Bernoulli naïve Bayes classifier is built with Boolean characteristics and executes the naïve Bayes training and classification methods for data distributed according to multivariate Bernoulli distributions [12].

The seventh model created was a Support Vector Machine Model, the objective of a Support Vector Machine Model is to evolve a model based on training data that forecasts the target values of the test data, given just the test data attributes [18]. The Support Vector Model generates the optimum line or decision boundary that can divide n- dimensional spaces into classes so that additional data points could be easily classified in the future [2]. Extra Tree Classifier was used to build the eighth model. The Extremely Randomized Trees Classifier is a form of ensemble learning approach that combines the results of several de-correlated decision trees gathered in a "forest" to get a classification result [8].

The ninth model was created using logistic regression. When the dependent variable is categorical, a logistic regression classifier is a frequently used machine learning method. It is the most repeatedly utilized algorithm for all classification tasks. Logistic Regression predicts the result by employing a sigmoid function, which returns a number between 0 and 1 [11]. The eleventh model was the K Nearest Neighbor Model (KNN). The KNN algorithm implies that comparable items exist in close vicinity, and in order to find the proper K for the data, the KNN method is run numerous times with distinct values of K, minimizing the number of errors and preserving the algorithm's capacity to generate correct predictions.

CatBoost algorithms were used to create the eleventh model. CatBoost algorithms build symmetric trees. [17] states that in every step, the preceding tree leaves are divided using the same requirement, and the feature-split pair with the lowest loss is chosen and utilized for all the level's nodes. The twelfth model was created using the Gradient Boost Classifier. The Gradient Boosting approach constructs an additive model in a forward stage-wise fashion; enabling for the evaluation of unspecified discrete loss functions [13].

Random Forest Classifier was used to generate the thirteenth model. A random forest is a meta figure that matches a range of call tree classifiers on several sub-samples of the dataset and utilizes averaging methods to improve predictive accuracy and manage overfitting [12]. The fourteenth model was created using the XGBoost Classifier. Extreme Gradient Boosting (XGBoost) is a robust, extended gradient-boosted decision tree machine learning algorithm [2]. It includes concurrent tree boosting and is the most popular machine learning algorithm for classifications, regression and ranking problems. The XGBoost method implementation includes multiple advanced capabilities for model adjustment, computing applications and algorithm developments, as well as the capacity to execute the three primary major types of gradient boosting [13].

The term "Passive-Aggressive" algorithms derives from the fact that if the prediction is accurate, the model can be retained and no modifications need to be made, while the term "Aggressive" refers to the necessity for changes in order to correct the model if the prediction is inaccurate [16]. Bagging Classifier was applied to the sixteenth model. An aggregate meta-estimator called a bagging classifier compares base classifiers on randomized portions of the original dataset, by merging their individual predictions through balloting or averaging to generate a final result. [4]. The seventeenth model was created using Linear Discriminant Analysis. A classifier that is intrinsically multi-class, generates closed-form solutions that are simple to compute, and has no parameters to dominate the learning procedure, is known as Linear Discriminant Analysis. By mapping the input data into a linear subdomain that contains the vectors that increase the difference between classes, Linear Discriminant Analysis can conduct supervised dimensionality reduction [3].

Quadratic Discriminant Analysis was used to build the eighteenth model. Quadratic Discriminant Analysis is a variant of Linear Discriminant Analysis classifiers and allows the non-linear separation of data. Quadratic Discriminant Analysis algorithms are less restrictive than Linear Discriminant Analysis classifiers and permit different feature covariance matrices for various classes, which results in a quadratic decision boundary [15].

The initial requirement to construct a machine learning model is a dataset because data is the foundation of all machine learning models. The dataset is data collection in a particular arrangement for a certain topic. In order to perform data pre-processing using Python [17], some predefined Python libraries needed to be imported and these libraries are used to perform certain specific jobs. The read csv() function of the pandas library is implemented to study a csv file and execute different actions on it to import the dataset. In machine learning, it is essential to determine the dataset's dependent and independent factors from the feature matrix. The Pandas library's iloc [] method is used to retrieve the necessary rows and columns from the dataset in order to obtain an independent variable [4].

The handling of missing values in the dataset is the next stage of data pre-processing. Missing data can be addressed in two key ways. Dealing with null values is the first method. In this manner, the particular row or column with zeros is eliminated. However, this method is ineffective because deleting data could lead to information loss, producing incorrect results [5]. Next, the average of the column or row holding any missing value is determined and used to fill in the gaps. A training set and a test set are created from the dataset. Using the test set, the machine learning model predicts the outcome. A training set is a portion of the dataset used to develop the machine learning model, and a testing set is a fraction of the dataset utilized to evaluate the machine learning model [17].

The model uses a confusion matrix to check the accuracy and the parameter confusion_matrix is imported from the sklearn.metrics library [8]. For each of the eighteen models created for the research, a colormap has been set for each confusion matrix created for each model, thereby enabling each model to have a unique colored confusion matrix [10]. For each of the eighteen models created for the research, a colormap has been set for each confusion matrix created for each model, thereby enabling each model to have a unique colored confusion matrix.

VI. IMPLEMENTATION

The implementation section is critical for obtaining the intended outcomes in the research. This chapter will provide a detailed review of how unsupervised and supervised learning algorithms produce optimum results like high accuracy scores, precision and recall values.

A confusion matrix is essential for machine learning because it analyzes how efficiently classification models perform when predictions are generated utilizing test data and indicates their suitability [18]. The type of errors made by the classifiers, which are either type-I or type-II errors, are also disclosed via a confusion matrix [13]. Additionally, the confusion matrix can be employed to calculate various factors such as the model's accuracy, precision, recall, and support values.

One of the crucial factors that affect how precisely classification problems are solved is classification accuracy [12]. It can be calculated as the ratio of the number of accurate predictions made by the classifiers and indicates how frequently the model predicts the appropriate output. The number of inaccurate guesses divided by the total number of predictions generated by the algorithm can be used to determine the error rate [15]. The number of accurate outputs supplied by the model or the proportion of correctly predicted positive classes by the model that actually occurred can be used to define precision [10]. Recall is the percentage of total positive classes for which the model accurately predicts the outcome. The highest possible recall value is required [12]. An F-score is utilized for this purpose since it is difficult to evaluate two models that have high recall values but low precision values. An F-score aids in simultaneously assessing the recall and precision values. If the recall and precision are equal, the F-score value is at its highest [14].

TABLE 1. SUMMARY OF THE ACCURACY SCORES, RECALLAND PRECISION VAUES OF ALL THE MODELS CREATED

ModelName	Accuracy Score (%)	Recall	Precision
Bernoulli Naïve Bayes	64%	0.64	0.64
Multinomial Naïve Bayes	65%	0.65	0.65
Decision Tree	86%	0.86	0.86
AdaBoost	51%	0.51	0.52
XGBoost	84%	0.84	0.84
Nu-Support Vector	86%	0.86	0.86
Extra Tree Classification	86%	0.85	0.86
Logistic Regression	65%	0.65	0.65
Passive-Aggressive	82%	0.82	0.82
Gradient Boosting	60%	0.60	0.61
K Nearest Neighbour	57%	0.57	0.57
Bagging Classification	83%	0.83	0.83
CatBoost	69%	0.69	0.69
Random Forest	85%	0.85	0.85
Linear Support Vector	76%	0.76	0.76
Quadratic Discriminant Analysis	83%	0.83	0.87
Linear Discriminant Analysis	79%	0.79	0.79
Support Vector Machine	84%	0.84	0.84

When there is a distinct line of separation between classes, the Nu-Support Vector Classification algorithm performs reasonably well, and it performs more efficiently in high-dimensional areas [5]. When there are more dimensions than samples, the Nu-Support Vector Classification algorithm performs skillfully and uses comparatively low memory [7]. Due to these reasons, Nu- Support Vector Classification algorithm obtained the highest accuracy of 86%. As shown in Table 1 above, Nu- Support Vector Classification model has the highest accuracy, precision, recall and f-1 value of 0.86, as this model has the lowest number of false positive and false negative values. Therefore, Nu-Support Vector Classification model is the best model created for this research.

The AdaBoost classifier was selected because it is simple to use, repeatedly corrects weak classifier errors, and increases accuracy by integrating weak classifiers [3]. However, because AdaBoost classifiers attempt to match each point precisely, they are sensitive to noisy data and are greatly impacted by outliers. [8]. In this research, the AdaBoost model obtained the lowest accuracy score as it has an extremely low average recall value of 0.51 as the AdaBoost model has more than 220 false negative values for each classification.

Even though Decision trees are classified as simple machine learning algorithms, they have several advantages such as interpretability which allows visualizing the decision tree and no pre-processing techniques are required to develop the data prior to creating the model [15]. A common factor of decision trees is data validity, as the algorithm effectively manages all the different kinds of data [13]. This is useful if the dataset has a combination of qualitative and quantitative data, and none of the categorial features need to be encoded. In this research, the Decision Tree model is the second- best modelobtained, as it has a high average recall value of 0.86 with a low number of false-negative values for all three classifications.

For this research, a fresh model for sentiment analysis was established based on the Recurrent Neural Network, as it is a state of-the-art technique in the Natural Language Processing field, but could not be achieved for the dataset as it is time-consuming. For instance, although libraries like Keras make the development of the neural network fairly simple, sometimes more control is needed over the details of the algorithm. Although modules like Tensorflow provide more opportunities, it is also more complicated, and the development takes much longer [9]. Furthermore, Recurrent Neural Network usually requires more customer review data compared to conventional machine learning algorithms, for example, at least thousands to millions of labeled samples. Moreover, as more data samples are required to build the model, the model based on the Recurrent Neural Network is more expensive when compared to traditional machine learning algorithms.

VII. CONCLUSION

For this research dissertation, from the comparative study of existing techniques that are available for sentiment analysis, machine learning methods were used to create the eighteen models to obtain the best model. To further improve the research, focus can be applied to combining machine learning algorithms with a lexicon-based approach to improve the precision of sentiment classification and apply sentimental analysis to various domains and languages.

A. *Contribution of Research*

This subsection will explain how this research can contribute to the industry. The ASOS company has a Twitter, Facebook, and Instagram as well as an official page. However, for this research, the reviews chosen for the dataset were retrieved only from the ASOS website and not from social media [10]. In this research, rather than segmenting the reviews based on age, gender, income and other surface demographics, each customer feedback on the ASOS website was retrieved and further segmented based on how the consumer actually feels about the brand and the products that they purchased on factors like customer experience, the condition of the product they obtained and delivery service.

This research shows that organizations are able to tackle negative feedback better and find a solution immediately when sentiment analysis is used tactically [14]. For example, for this research most negative sentiments on the ASOS website were from people who did not like the delivery process. Hence, ASOS can make an operational change by replacing their deliverypartner Hermes with a new delivery company like DHLto improve their sales and satisfy their customers.

Furthermore, this research shows how sentiment analysis facilitates the growth of an organization by implementing the changes based on the customers feeling toward a particular product or brand. [3]. For instance, customer sentiment increased after April 2020, as the ASOS website had introduceda new chat session that enabled customers to directly type and send queries using the chat option, rather than calling the customer service hotline [5]. This change made at the business enabled to boost more positive reviews,as more people were able to share their concerns and receive feedback.

B. Future Work

This research can be upgraded by choosing the reviews available on the ASOS social media sites, as every time the major social media platforms update themselves and add additional options, the information behind those interactions gets broader and deeper. Over recent years,there have been massive leaps in machine learning and artificial intelligence, and many analytical solutions are looking to replace these technologies; to replace algorithms[18].

Unfortunately, the current data provided on the ASOS website does not allow to measure the audience emotions as machine learning is not yet ready to tackle the complex nuances of text and the complicated communication methods used by several individuals, especially on social media channels that are rife with slang, sarcasm, double meanings and misspellings. Hence, it is difficult for artificial intelligence systems to accurately sort and classify sentiments on the websites and social media thereby obtaining very low precision, recall and accuracy values for the research.

Thus, it can be concluded that to improve the research for future work, the audience's emotions and images should be analyzed and not only the text given in the customer feedback. Interpreting social media data will provide more insights for the researcher compared to the customer reviews provided on the ASOS website as social media data contain emoticons and different kinds of language which actually emphasize what a user feels like.

REFERENCES

[1] B. Pavlyshenko, "Using machine learning and semantic features in intellectual analysis of text data," Electronics and Information Technologies, vol. 13, 2020, doi: 10.30970/eli.13.1.

[2] Z. Shaikh, A. Mohadikar, R. Nayak, and R. Padamadan, "Security and Verification of Server Data Using Frequent Itemset Mining in Ecommerce," International Journal of Synthetic Emotions, vol. 8, no. 1, pp. 31–43, Jan. 2017, doi: 10.4018/ijse.2017010103.

[3] Dr. S. V., "Text Analysis Via Composite Feature Extraction," Journal of Advanced Research in Dynamical andControl Systems, vol. 24, no. 4, pp. 310–320, Mar. 2020, doi:10.5373/jardcs/v12i4/20201445.

[4] J. XIANG, "Human motion data analysis and retrievalbased on 3D feature extraction," Journal of Computer Applications, vol. 28, no. 5, pp. 1344–1346, May 2008, doi: 10.3724/sp.j.1087.2008.01344.

[5] Y. Sun and D. Wu, "Feature extraction through local learning," Statistical Analysis and Data Mining, vol. 2, no.1,pp. 34–47, Jun. 2009, doi: 10.1002/sam.10028.

[6] D. Chien, M. Featherstone, K. Min, and C. Rogers, "Speech separation with artifical neural networks," Neural Networks, vol. 1, p. 294, Jan. 1988, doi: 10.1016/0893-6080(88)90325-5.

[7] P. Géczy and S. Usui, "Rule Extraction from Trained Artifical Neural Networks," Behaviormetrika, vol. 26, no. 1,pp. 89–106, Jan. 1999, doi: 10.2333/bhmk.26.89.

[8] A. Botchkarev, "Evaluating Performance of Regression Machine Learning Models Using Multiple Error Metrics in Azure Machine Learning Studio," SSRN Electronic Journal, 2018, doi: 10.2139/ssrn.3177507.

[9] O. Harfoushi and D. Hasan, "Amazon Machine Learning vs. Microsoft Azure Machine Learning as Platforms for Sentiment Analysis," International Journal of Advanced Science and Technology, vol. 118, pp. 131–142,Sep. 2018, doi: 10.14257/ijast.2018.118.12.

[10] O. Harfoushi, D. Hasan, and R. Obiedat, "Sentiment Analysis Algorithms through Azure Machine Learning: Analysis and Comparison," Modern Applied Science, vol.12, no. 7, p. 49, Jun. 2018, doi: 10.5539/mas.v12n7p49.

[11] H. M and Dr. B. M. Sagar, "Smart Health Care Implementation Using Naïve Bayes Algorithm," International Journal of Innovative Research in ComputerScience & Technology, vol. 7, no. 3, pp. 90–93, May 2019, doi: 10.21276/ijircst.2019.7.3.11.

[12] G. Yan and X. Huang, "Improved harmonic-balancealgorithms based on the genetic algorithms," JOURNAL OFELECTRONIC MEASUREMENT AND INSTRUMENT,vol. 2009, no. 10, pp. 96–99, Dec. 2009, doi: 10.3724/sp.j.1187.2009.10096.

[13] F. F. Balahadia and B. E. V. Comendador, "Adoption ofOpinion Mining in the Faculty Performance Evaluation System by the Students Using Naïve Bayes Algorithm,"International Journal of Computer Theory and Engineering, vol. 8, no. 3, pp. 255–259, Jun. 2016, doi: 10.7763/ijcte.2016.v8.1054.

[14] C. J. Mantas, J. G. Castellano, S. Moral-García, and J. Abellán, "A comparison of random forest based algorithms: random credal random forest versus oblique random forest,"Soft Computing, Nov. 2018, doi: 10.1007/s00500-018-3628-5.

[15] S. Jarvis, N. A. Dimarzio, R. P. Morrissey, J. Ward, and Moretti, "Automated classification of odontocetes in open ocean environments using a novel multi-class suport vector machine," The Journal of the Acoustical Society of America,vol. 123, no. 5, pp. 2984–2984, May 2008, doi: 10.1121/1.2932507.

[16] R. Karimov, "Combined Machine-Learning Approach to PoS-Tagging of Middle English Corpora," Crossroads. A Journal of English Studies, no. 21(2), pp. 42–52, 2018, doi: 10.15290/cr.2018.21.2.04.

[17] M. Meelen and N. Hill, "Segmenting and POS tagging Classical Tibetan using a memory-based tagger," HimalayanLinguistics, vol. 16, no. 2, Jan. 2018, doi: 10.5070/h916234501.

[18] M. Prajapati and A. Yajnik, "POS Tagging of Gujarati Text using VITERBI and SVM," International Journal of Computer Applications, vol. 181, no. 43, pp. 32–35, Mar.2019, doi: 10.5120/ijca2019918514.