YOUCEFI, M.R., BOUKREDERA, F.S., GHALEM, K., HADJADJ, A. and EZENKWU, C.P. 2024. Development of an expert-informed rig state classifier using naive Bayes algorithm for invisible loss time measurement. *Applied intelligence* [online], 54(17-18), pages 7659-7673. Available from: https://doi.org/10.1007/s10489-024-05560-5

Development of an expert-informed rig state classifier using naive Bayes algorithm for invisible loss time measurement.

YOUCEFI, M.R., BOUKREDERA, F.S., GHALEM, K., HADJADJ, A. and EZENKWU, C.P.

2024

Development of an Expert-Informed Rig State Classifier Using Naïve Bayes Algorithm for Invisible Loss Time Measurement

Mohamed Riad Youcefi [a, c, *], Farouk Said Boukredera [b], Khaled Ghalem [a], Ahmed Hadjadj [c], Chinedu Pascal Ezenkwu [d]

[a] Department of Process Engineering, Faculty of Technology, University Amar Telidji-Laghouat, BP37G 03000 Laghouat, Algeria

[b] Centre for Applied Dynamics Research, University of Aberdeen, UK

[c] Laboratory of Petroleum Equipment's Reliability and Materials, Faculty of Hydrocarbons and Chemistry, University M'hamed Bougara of Boumerdes, Avenue de l'Indépendance, 35000 Boumerdes, Algeria

[d] Business Analytics, School of Creative and Cultural Business, Robert Gordon University, UK

* Corresponding Author: Email: mr.youcefi@lagh-univ.dz, m.youcefi@univ-boumerdes.dz; Phone : +213699434802

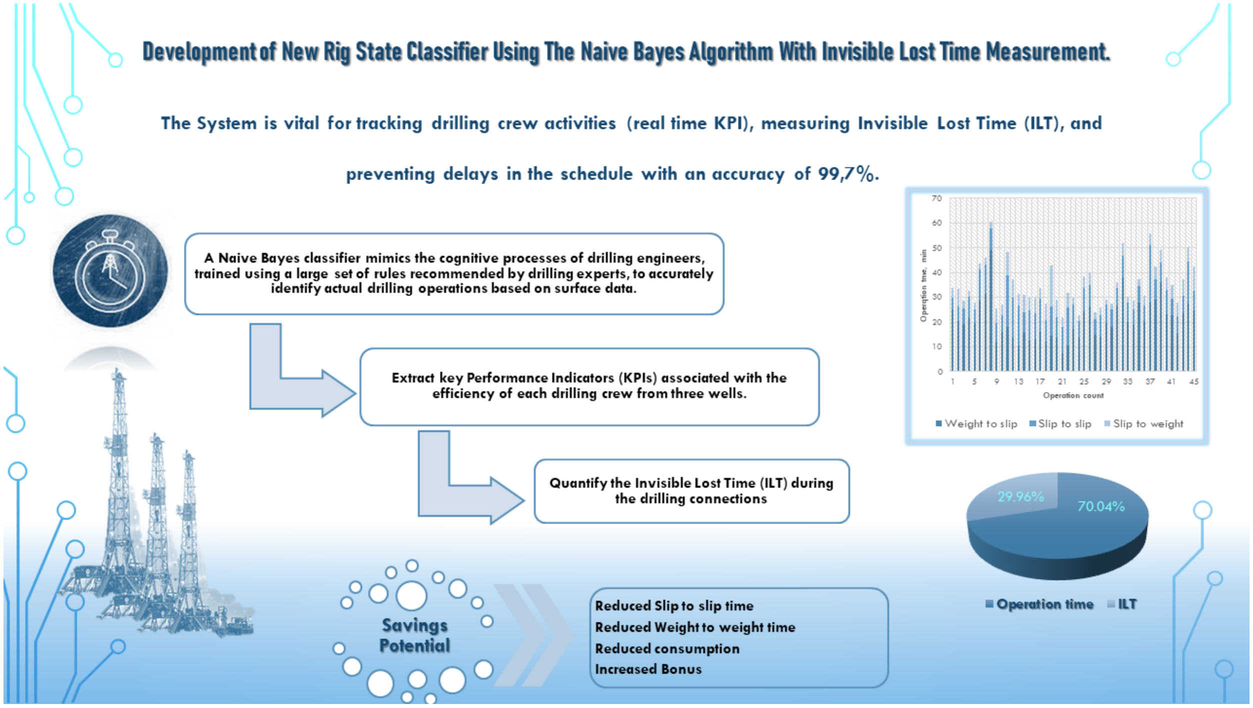
Development of an Expert-Informed Rig State Classifier Using Naïve Bayes Algorithm for Invisible Loss Time Measurement.

Abstract:

The rig state plays a crucial role in recognizing the operations carried out by the drilling crew and quantifying Invisible Lost Time (ILT). This lost time, often challenging to assess and report manually in daily reports, results in delays to the scheduled timeline. In this paper, the Naive Bayes algorithm was used to establish a new trustworthy rig state. Training data, consisting of a large set of rules, was generated based on drilling experts' recommendations. This dataset was then employed to build a Naive Bayes classifier capable of emulating the cognitive processes of skilled drilling engineers and accurately recognizing the actual drilling operation from surface data. The developed model was used to process high-frequency drilling data collected from three wells, aiming to derive the Key Performance Indicators (KPIs) related to each drilling crew's efficiency and quantify the ILT during the drilling connections. The obtained results revealed that the established rig state excelled in automatically recognizing drilling operations, achieving a high success rate of 99.747%. The findings of this study offer valuable insights for drillers and rig supervisors, enabling real-time visual assessment of efficiency and prompt intervention to reduce ILT.

Keywords: Rig state; Drilling; Machine learning; Nave Bayes classifier; Invisible Lost Time; key Performance Indicator.

Graphical abstract


Development of New Rig State Classifier Using The Naive Bayes Algorithm With Invisible Lost Time Measurement.
The System is vital for tracking drilling crew activities (real time KPI), measuring Invisible Lost Time (ILT), and preventing delays in the schedule with an accuracy of 99,7%.
A Naive Bayes classifier mimics the cognitive processes of drilling engineers, trained using a large set of rules recommended by drilling experts, to accurately identify actual drilling operations based on surface data.
Extract key Performance Indicators (KPIs) associated with the efficiency of each drilling crew from three wells.
Quantify the Invisible Lost Time (ILT) during the drilling connections
Savings Potential
Reduced Slip to slip time
Reduced Weight to weight time
Reduced consumption
Increased Bonus
Operation count
Weight to slip
Slip to slip
Slip to weight
29.96%
70.04%
Operation time
ILT

1. Introduction

The drilling process is one of the most important, complicated, and expensive procedures within oil and gas development projects [1, 2]. Therefore, it is critical to optimize the return on investment in drilling operations, particularly during periods of low oil prices. According to project management standards, drilling operations must consistently meet deadlines and stay within budget constraints. However, challenges such as drilling issues and undesirable events frequently induce delays, causing the drilling operation to fall behind schedule. This period spent dealing with problems related to drilling activities, such as stuck pipes, waiting for materials, or weather-related interruptions, is defined as Non-Productive Time (NPT) [3, 4].

NPT can be broadly described as any time not spent constructing the well [5]. NPT consists of visible and Invisible Lost Time (ILT). ILT, specifically, refers to time lost during common operations such as tripping and connections, where maximum efficiency is impeded by factors like insufficient experience, proficiency, or the absence of appropriate tools [6, 7]. Drilling multiple wells with the same rigs under similar conditions indicated that the time required to complete certain activities differs from well to well. This divergence is attributed to the fact that only visible NPT is assessed and reported in daily reports, while time lost due to process inefficiencies remains unaccounted for.

Despite these inefficiencies translating to only a few seconds or minutes of additional operating time per operation, their cumulative effect can lead to a substantial amount of wasted time over the entire working timeline. De Oliveira et al. [8] stated that the visible NPT can be measured manually, while the ILT assessment necessitates an automated system due to the large amount of real-time data generated, making manual measurement impractical at the well site or monitoring center.

Machine Learning (ML) algorithms have gained substantial popularity as excellent alternatives for studying complex systems [9–12]. These techniques play an increasingly vital role in addressing challenges related to high-complexity computation and reliability across various fields of study. Various researchers have witnessed the practical implementation of intelligent modeling approaches in the oil and gas field. Notably, in drilling engineering, researchers [13–16] have successfully applied various ML algorithms to classify the severity of stick slips and optimize the drilling process. Furthermore, numerous methodologies and approaches, based on comprehensive drilling data analysis and machine learning algorithms, have been carried out to quantify ILT and help assist operators in managing the drilling operation. The majority of ILT measurement methodologies, as reported in the literature, rely on the rig state, which represents an intelligent system that can ensure continuous real-time identification of a rig's activities.

In this paper, a novel robust rig state was developed using the Naive Bayes algorithm. Distinguishing itself from prior studies that used surface data for rig state training, our primary contribution lies in employing expert-crafted rules. This results in high-quality, trustworthy data for training the rig state classifier to mimic the thought process of a drilling expert.The developed Naive Bayes classifier model was applied to process and label high-frequency time series data collected from the surface sensors across three wells. This analysis aimed to assess the performance of many drilling crews and derive the KPIs related to each drilling crew's efficiency. Subsequently, ILT was derived and graphically plotted to highlight the real-time practical application of the rig state.

The rest of the paper is organized as follows: Section 2 delves into related works and outlines the contributions of the current study. Section 3 provides a detailed explanation of the database employed in this research. Section 4 explains the workflow involved in the rig state development. Results from the evaluation and application of the established rig state are presented and discussed in Sections 5 and 6. Finally, Section 7 highlights the main outcomes derived from the study.

2. Background and literature survey

The current section is dedicated to discussing the existing research on the enhancement of drilling performance through the application of machine learning and other strategies. A significant focus in many of these studies lies in the real-time implementation of ML algorithms to detect undesirable drilling events and mitigate the effect of these events, which can dramatically increase NPT. Muojeke et al. [17] proposed an approach that integrates Artificial Neural Network (ANN) and a binary classifier for early kick detection based on downhole parameters. Gurina et al. [18] applied ML to forecast the occurrence of six types of drilling accidents. Tran et al. [19] introduced a framework that combines the Convolutional Neural Network (CNN) with Long-Short-Term Memory (LSTM) for detecting drill bit failures. Furthermore, Mopuri et al. [20] presented a novel approach for the early detection of stuck pipe events using deep learning. While these works have made a significant effort in mitigating the unwanted drilling events, it is essential to recognize they only address the visible lost time.

Numerous endeavors have been undertaken to establish a foundation for in-depth analysis of ILT. De Oliveira et al. [8] provided examples of offshore drilling improvements achieved through the automation of the rig state detection technique. Using the automated classification of drilling operations, they accurately identified reasons for certain rigs not meeting contractual objectives. Lakhanpal and Samuel [21] proposed a method to measure ILT by analyzing intrinsic mode functions of drilling data using empirical mode decomposition. Additionally, Al Ady et al. [22] conducted a performance analysis of the entire rig fleet to derive the benefits of monitoring the drilling through a set of examined KPIs. Their study highlighted the need for a system capable of segregating drilling activities, revealing the significance of developing a rig state model for similar tasks. However, it is crucial to note that these studies are primarily regarded as post-analyses of the drilling performance to extract insights related to the underperforming areas.

Recognizing the imperative for a robust real-time system to enhance operational efficiency, several methodologies have been proposed to develop a classification model capable of automatically recognizing drilling operations through strategic application of ML and statistical approaches. Zhao et al. [23] devised an approach for automatic slip status and stand detection, enabling the computation of KPIs. Although this proposed approach is effective for vertical, horizontal, and extended-reach drilling wells, it might not be suitable for inclined wells. Coley [24] implemented various supervised ML algorithms to build a classifier identifying rig states from surface data. Ben et al. [25] applied Random Forest, CNN, and a hybrid Convolutional Neural Network/Recurrent Neural Network (CNN/RNN) to label high-frequency time-series data with various rig states The output from their rig-state classifier contributed to generating KPI data for supporting ILT Analysis, resulting in significant reductions in connection times. Yin et al. [26] utilized ANN to develop a rig state classification model, enabling real-time identification of rig states and subsequent performance evaluation against KPIs for ILT detection. Additional ML applications for rig state classification were proposed by Arnaout et al. [27] to detect common drilling operations based on polynomial approximation..

Despite the high accuracy achieved by the classification models developed in the aforementioned studies, it is crucial to note that these models were trained using field data collected from surface rig sensors. This surface data is susceptible to noise and lacks the complete insights required to distinguish between different rig states. This limitation makes rig state models trained with field data prone to inaccuracies in recognizing drilling operations, leading to imprecise estimations of KPIs and ILT.

2.1. Motivation and contribution

In contrast to prior studies relying on field data for rig state model training, the present study employs trustworthy data that provides comprehensive information about rig state identification. The dataset introduced in this paper for training the rig state classifier was carefully curated by drilling experts, and it consists of a set of rules followed by skilled drilling engineers to identify rig states using surface data. This strategic choice in training the rig state classifier aims to achieve a higher success rate in the classification of drilling operations.

Moreover, in this study, the Naive Bayes classifier was employed for the first time, to the best of our knowledge, to construct a rig state model, leveraging drilling experts' recommendations. The Naive Bayes classifier, known for its simplicity, ease of implementation, and high performance, was chosen due to its proficiency in handling categorical data, often surpassing more complex classification approaches [28–31].

By leveraging the efficiency of Naïve Bayes in handling categorical data and employing trustworthy training data, the current study aims to build a robust rig state model. This model is intended for real-time implementation to accurately recognize drilling operations and seamlessly integrate into the monitoring system for deriving KPIs and measuring ILT. The subsequent section provides an in-depth explanation of the data insights, characteristics, and sources.

3. Database generation for Naive Bayes Classifier training

The quality of the training data plays a crucial role in developing a robust ML model. Training the rig state classifier exclusively with field data may result in a classification model that struggles to identify the rig state. This limitation arises from the potential noise in the surface data recorded by rig sensors. Additionally, the surface data might lack comprehensive insights needed for effective discrimination between various rig states. The paramount contribution of this paper lies in the use of reliable data for the development of the rig state. This dataset was generated based on the recommendations of drilling experts. It consists of complete information and rules that a skilled drilling engineer adheres to when discerning drilling operations through surface data. The employment of such trustworthy data holds the potential to create an intelligent system classifier that emulates the cognitive processes employed by skilled drilling engineers in the classification of drilling operations.

Table 1 shows a sample of data generated for training a drilling operation classification model. The first six columns in **Table 1** correspond to input variables for the training data, including Drill bit Position, Weight on Hook (WOH), Traveling Block Motion, Revolutions Per Minute (RPM), Drilling Torque, and Flow Rate. Each input vector consists of various attributes associated with the respective input variable. For instance, the WOH variable has the following attributes: "Low" if the weight of the hook equals the weight of the traveling block and its accessories, or "High" if the weight of the hook surpasses the weight of the traveling block and its accessories.

The variable 'Drill bit Position' can assume the following attributes: "Null" when the difference between the depth and the drill bit position is zero, and "Positive" when the drill bit is not positioned at the bottom of the well. The traveling block's motion is described by three attributes based on the difference between the current position of the traveling block P_2 and the previous position of the traveling block P_1: "Up" if the difference between P_2 and P_1 is positive, "Down" if the difference between P_2 and P_1 is negative, and "Fixed" if the difference between P_2 and P_1 is zero. The remaining input variables, RPM, torque, and flow rate are represented by two attributes: "Positive" if the value exceeds a predefined threshold, and "Null" otherwise.

The last column corresponds to the vector of the output variable, which includes various classes representing drilling operations such as drilling, back reaming, reaming, in slips, circulation, Running In Hole (RIH), and Pulling Of Out Hole (POOH), or the static state (stationary).

Table 1 Data sample for training the Naïve Bayes Classifier

WOH	Drill bit Position	Traveling Block Motion	RPM	Drilling torque	Flow rate	Rig state
High	Null	Down	Positive	Positive	Positive	Drilling
High	Positive	Down	Null	Null	Null	RIH
Low	Positive	Up	Null	Null	Null	Inslips
Low	Positive	Down	Null	Null	Null	Inslips
Low	Positive	Fixed	Null	Null	Null	Inslips
High	Positive	Up	Null	Null	Null	POOH
High	Positive	Up	Positive	Positive	Positive	Back reaming
High	Positive	Down	Positive	Positive	Positive	Reaming
High	Positive	Fixed	Null	Null	Positive	Circulation
High	Positive	Fixed	Null	Null	Null	Stationary

It is worth noting that in some cases, sensors may provide positive values (low values) for torque, RPM and flow even when there is no longer rotation, and the pumps are turned off. To address these potential sensor malfunctions, which can occur at any moment, threshold values are set at 15 rpm, 1000 ft.lb, and 100 l/min for RPM, torque, and flow rate, respectively.

4. Workflow of Rig State Model Development

4.1. Training the Naive Bayes Classifier

In this section, the Naive Bayes algorithm was applied to create a robust classifier capable of accurately detecting and recognizing the rig state in real time based on surface data. The training of the Naive Bayes classifier involves determining the prior probabilities $P(C_j)$ and the likelihood $P(X_i/C_j)$ from the data presented in **Table 1**. **Fig. 1** illustrates the computation of probabilities $P(X_i/C_j)$ for the variable "Traveling Block Motion". These probabilities measure the possibility of having downward, upward, or no motion of the traveling block when the operation type is known.

As depicted in **Fig. 1**, the likelihood of the traveling block moving upward when the actual operation is drilling is zero. Furthermore, the possibility of the traveling block moving during the stationary state is also zero. These training results demonstrate the capability of the built Naive Bayes classifier to emulate the reasoning of engineers when attempting to determine the drilling operation from drilling data. Based on these learning results, it is anticipated that the classifier developed in this study will be able to identify drilling operations with a high success rate. Similarly, the probabilities $P(X_i/C_j)$ for the other variables were calculated from the training data.

Operation		P(operation)
Drilling	2	2/23
RIH	1	1/23
POOH	2	2/23
In slips	7	7/23
Back reaming	2	2/23
Reaming	2	2/23
Circulation	4	4/23
Stationary	3	3/23
Total	23	1

Operation	Down	Fixed	Up	Total
Drilling	1	1	0	2
RIH	1	0	0	1
POOH	0	0	1	1
In slips	1	5	1	7
Back reaming	2	0	0	2
Reaming	0	0	2	2
Circulation	0	4	0	4
Stationary	0	3	0	3

Operation	P(down/operation)	P(Fixed/operation)	P(Up/operation)
Drilling	1/2	1/2	0
RIH	1	0	0
POOH	0	0	1
In slips	1/7	5/7	1
Back reaming	1	0	0
Reaming	0	0	1
Circulation	0	1	0
Stationary	0	1	0

Fig. 1 The computation of the probabilities $P(X_i/C_j)$ for the variable " Traveling Block Motion".

4.2. Implementation of the developed rig state- Naive Bayes classifier

Following the determination of prior probabilities, the created Naive Bayes model was employed to predict drilling operations based on surface drilling data. However, real-time application of the established Naive Bayes classifier necessitates converting surface drilling data from numerical to categorical variables. **Fig. 2** illustrates the process used in this study to convert drilling data into categorical variables before predicting the rig state.

As depicted in **Fig. 2**, this data transformation is executed through a conditional structure stated in several blocks of code. The first block receives data from sensors, intermediate blocks process the data to extract insights, and the last block assigns an attribute to each input variable based on the conditional expressions in the if and else statements.

Following data processing, the posterior probability $P(C_j/X_i)$ of each class C_j (e.g., the probability of detecting a back reaming operation noted C_j knowing that the flow rate X_i is positive) is estimated using Bayesian independence theories between the variables. The Bayesian Naive classification algorithm then searches through all the drilling operations to find the one with the highest posterior probability, as described in the following equation:

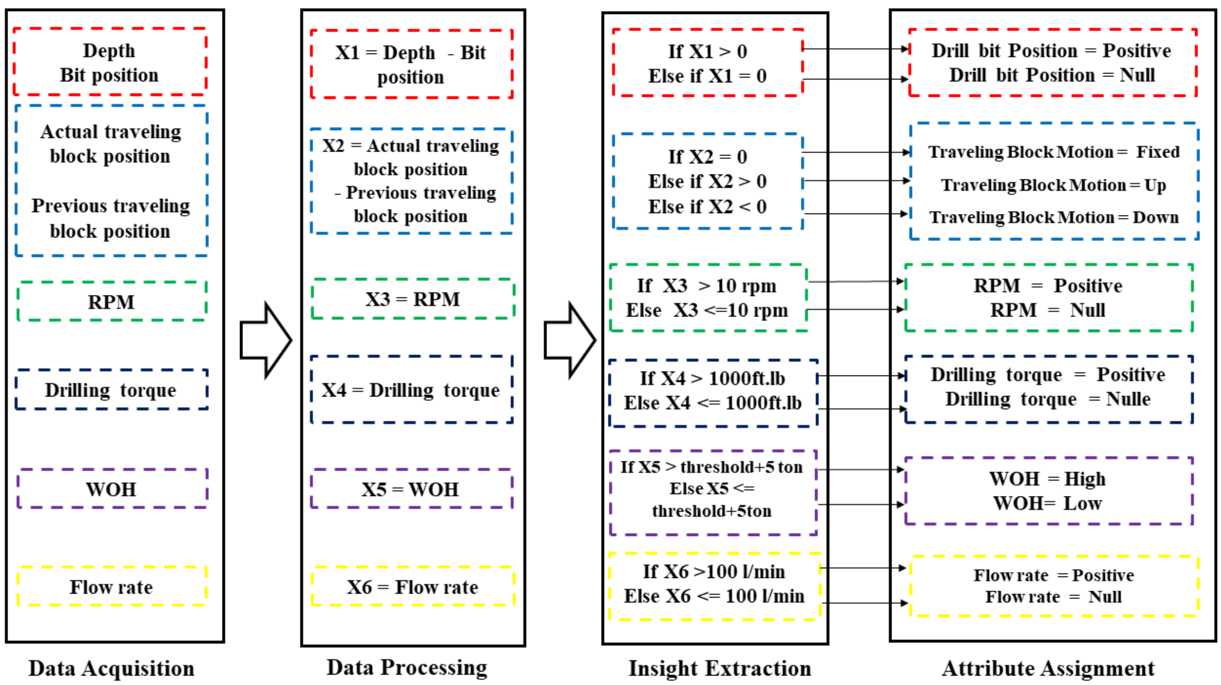


Fig. 2 The procedure used for converting surface drilling data from numeric variables to categorical variables.

$$M = \text{Argmax} \frac{\prod_i P(x_i \mid c_j)P(c_j)}{\sum_j^k \prod_i P(x_i \mid c_j)P(c_j)} \quad (1)$$

where M is the operation predicted by the Naive Bayes classifier and k is the number of drilling operations k=8.

5. Model Evaluation: Rig State Performance Assessment

In this section, the developed Naive Bayes classifier was evaluated using drilling data gathered from surface sensors during drilling the section $12^{1/4"}$ of well-A to test its ability to predict the rig state in real time. The Naive Bayes model's performance was assessed using two metrics: Accuracy (Acc) and precision (P). The accuracy metric measures how the model performs in general across all classes. It is calculated by dividing the number of accurate predictions by the total number of predictions. The precision metric describes the model's performance within a single class C_j. It is computed by dividing the number of well-classified operations in class C_j by the total number of operations categorized in class C_j (whether correct or incorrect).

$$\text{Accuracy } = \frac{\text{number of correct predictions}}{\text{total number of predictions}} \tag{2}$$

$$\text{Precision } = \frac{\text{number of well classified operations in class } C_j}{\text{total number of operations classified in } C_j} \tag{3}$$

The performance achieved by the Naive Bayes model is highlighted in **Table 2**. The testing results revealed that the developed Naive Bayes classifier was able to accurately classify the drilling operations with a 99.747% accuracy rate. There were instances where the model misclassified them as reaming operations instead of circulation operations. A closer examination of the misclassified cases revealed that they involved data points with null torque and rpm values, the traveling block position indicating a downward movement, and the created model encountered difficulties in distinguishing between reaming and circulation operations, resulting in the misclassification. It is worth noting that these instances constituted only 0.253% of the whole testing data, underscoring the overall exceptional high performance of the developed rig state model, with accurate classification achieved in almost all cases.

Fig. 3 depicts the timeline of the automated drilling operation recognition against drilling data. The established rig state accurately detected the drilling operation based on surface data. For example, when the WOH was reduced, the system successfully recognized the rig state 'in slips', as illustrated by the blue line.

Table 2 The performance of the established Rig State model

Rig state	Total number of predictions	Number of accurate predictions	Precision (%)
Drilling	31175	31175	100
Back reaming	9405	9405	100
Reaming	5008	5008	100
Stationary	1606	1606	100
Circulation	11191	11016	98.436
In slips	8098	8098	100
POOH	2091	2091	100
RIH	540	540	100
Overall	Total number of predictions	Number of accurate predictions	Accuracy (%)
	69114	68939	99.747

The developed Naive Bayes classifier was compared with other ML algorithms such as the Support Vector Machine (SVM) and the Decision Tree (DT) algorithm. The training data described in Section 2 was used to train the SVM and the DT to build additional rig state classifiers. Subsequently, data from well A was used to compare the performance of the three models (NB, SVM, and DT).

To ensure fair competition among the compared ML models, hyperparameter tuning was performed for each model to achieve the best configuration and structure of all the models before comparing them. The tuning process involved a trial and error approach, optimizing parameters such as the regularization parameter (C), kernel function, kernel coefficient for SVM, and parameters including the number of trees, maximum depth, minimum samples split, and minimum samples leaf for the DT algorithm. By executing multiple runs and exploring different hyperparameter values, the best combination that presented highest accuracy was identified. The performance obtained by each model is highlighted in **Table 3**.

Here, he DT algorithm was comparatively less accurate with a success rate of 99.6411%. NB and SVM achieved accuracy rates of 99.7467% and 99.7395%, respectively. The performance of both SVM and NB showed similarity, with SVM failing to correctly identify five instances compared to NB. Besides being a less complex classifier, the NB exhibited good performance compared to the other ML algorithms.

Table 3 A comparison between the ML models.

Model	NB	SVM	DT
Total number of predictions	69114	69114	69114
Number of accurate predictions	68939	68934	68866
Accuracy (%)	99.7467	99.7395	99.6411

6. Case Study : Rig State Application

The present section aims to highlight the significance of the developed rig state model along with its practical applications and benefits in the drilling field. The built Naive Bayes classifier was implemented in a comprehensive case study to evaluate its effectiveness in accurately labeling large drilling data collected from three wells. The primary objective was to extract valuable insights, specifically the KPIs and the ILT. These insights can provide drilling engineers with enhanced knowledge of drilling efficiency and enable real-time monitoring.

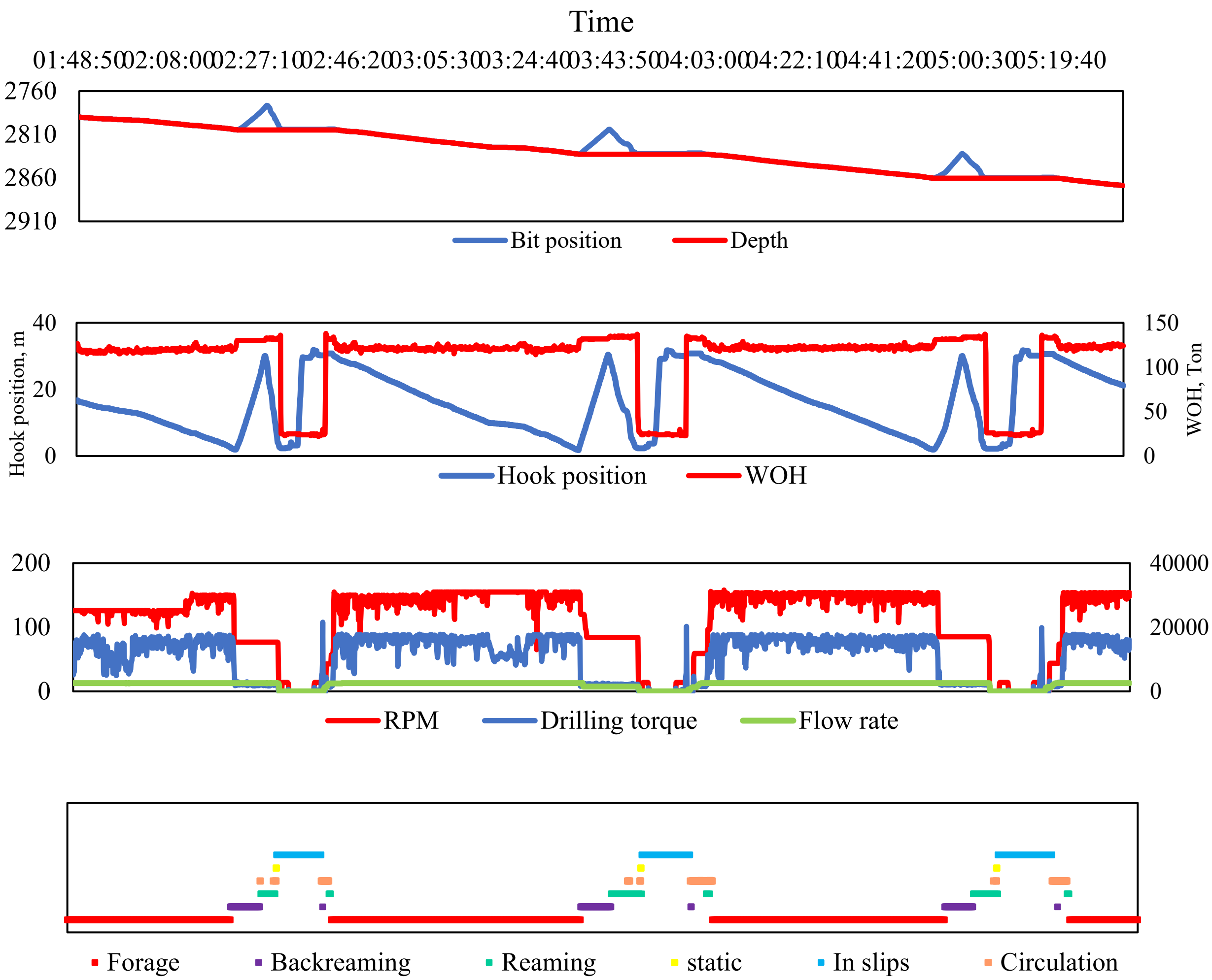


Fig. 3 *The rig state identified using the established Naïve Bayes classifier vs the corresponding surface data*

6.1. Key performance indicator assessment

ILT is measured by examining the individual KPI generated by a drilling crew, equipment, or a combination of both. The focus of this case study was on connection time during drilling operations for three wells and six crews in an Algerian field. The built rig state was used to process and label high-frequency time series data collected from surface sensors during the drilling of the 16", $12^{1/4"}$ and the 8½" diameter sections. The main objectives were to identify the drilling activity, quantify the time each crew spent on a specific routine drilling operations, and derive KPIs for connection time.

The crew's performance during connection time was assessed in this study through four distinct KPIs: namely weight to weight, weight to slip, slip to slip, and slip to weight. During the drilling connection, if the formations appeared to have issues with swelling or other issues that increase the risk of getting stuck, the driller could ream and back ream the formation to clean out the hole before the crew set the slips to grip the drill string [32]. Depending on when they are performed, the time spent on reaming, back reaming, circulation, and the well survey is considered as part of the weight to slip or slip to weight time.

- Weight to slip is the time spent putting the drill string into the slips after drilling a stand. This time can include the following rig states: stationary, reaming, back reaming, and circulation.
- Slip to weight is the time spent placing the drill bit on the bottom after the drill string has been taken out of slips. This time can include the following rig states: stationary, reaming, back reaming, and circulation.
- Slip to slip is the time spent between putting the drill string into slips and taking it out of slips. This time can include the following rig states: in slips.

- Weight to weight is the time spent picking up the drill bit from the bottom after drilling a stand and setting it back on the bottom after making a connection. It represents the sum of the three aforementioned KPIs (weight to slip, slip to slip, and slip to weight).

Fig. 4 shows how the four KPIs are derived using the proposed rig state model. The developed rig state processed and labeled high-frequency time series data collected from surface sensors. The rig states such as reaming, back reaming, static, and circulation, which are detected after drilling a stand and before putting the drill string into the slips were merged inside the weight to slip block by the system, and the corresponding KPI was computed. Similarly, the rig states such as reaming, back reaming, static, and the circulation that are recognized after taking the drill string out of slips until starting to drill a new stand were automatically combined by the system inside the slip to weight block and the relevant KPI was computed. Only the rig state in slips was considered for computing the "slip to slip" KPI.

Finally, the "weight to weight" KPI was calculated by summing up the three KPIs: weight to slip, slip to slip, and slip to weight. Following the computation of the KPIs, separate histograms for each crew were prepared and displayed in **Fig. 5** to compare various KPIs between wells and crews.

The “weight to weight” time represents the most significant duration performed between putting the drill string into the slips and repositioning it on the bottom hole after making a connection. This is due to the necessity of conducting multiple reaming and back-reaming trips to avoid getting stuck. Although the crews in the three wells were working under similar rig performance and conditions, the results revealed a notable disparity in performance during connections. The night crew of well-B performed this operation routinely faster than the other crews, with an average "weight to weight" time of 31 minutes. The other crews performed poorly compared to the night crew of well-B, with their average "weight to weight" time ranging between 34 and 43 minutes. The derived "weight to weight" time for drill pipe connections in well C shows that the shortest "weight to weight" time is 13 minutes (day crew, standpipe number: 5).

These results highlight the importance of integrating the proposed rig state into the real-time monitoring process. For example, the poor performance of the day crew observed in well A could be promptly identified, diagnosed for its underlying causes, and thereby contribute to an enhancement in operational efficiency.

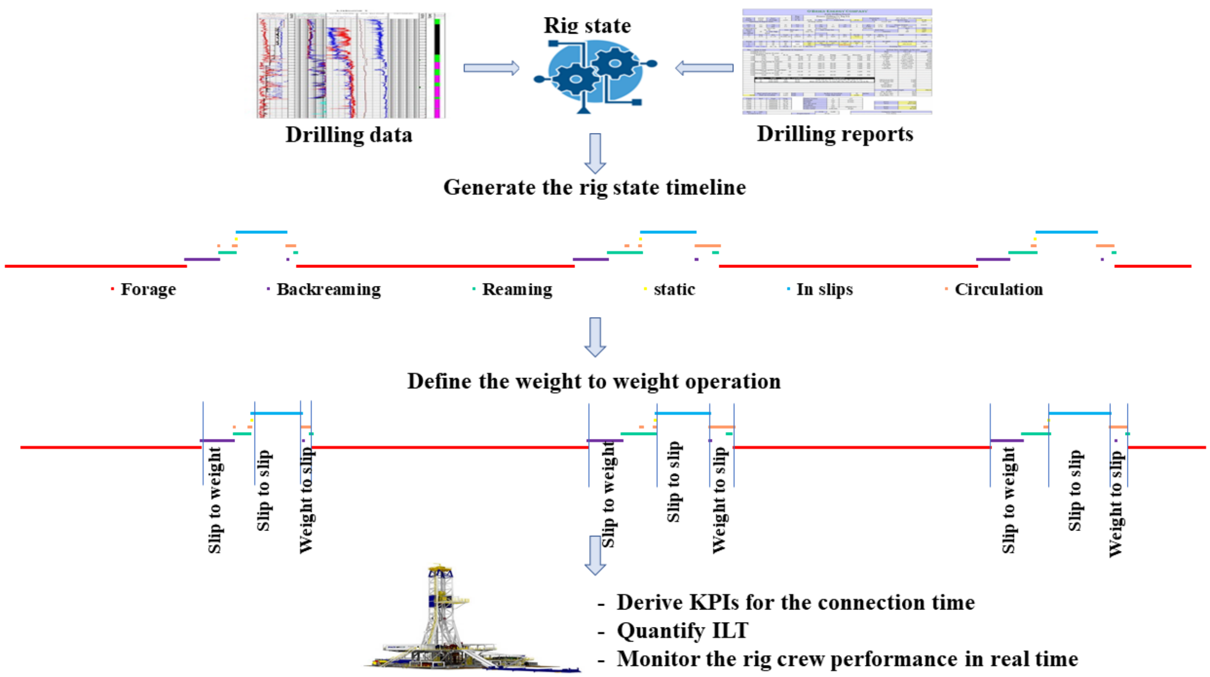


Fig. 4 KPIs determination using the rig state

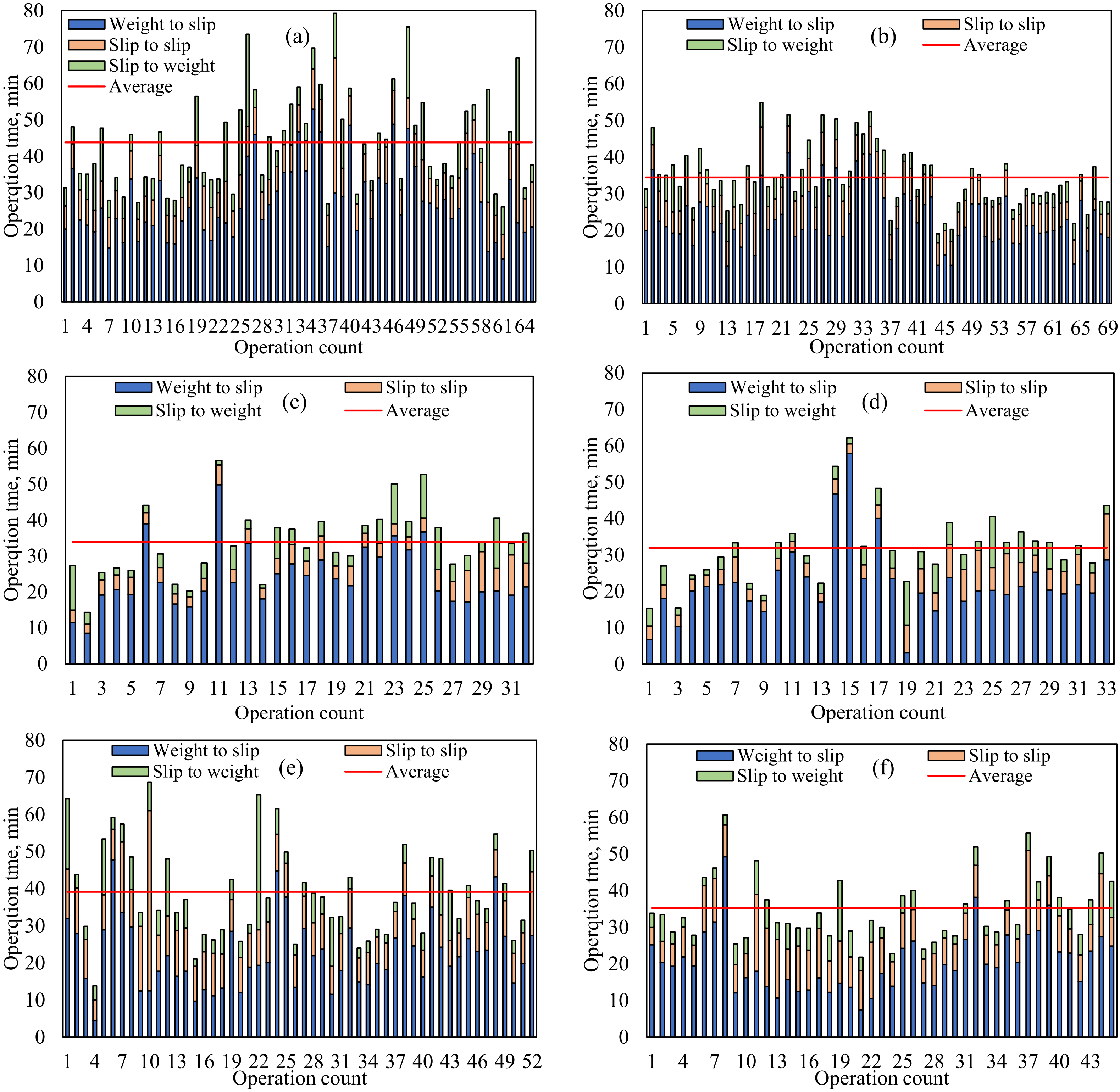


Fig. 5 *Weight to Weight time Comparison of all crews in the three wells : (a) Day crew in well A; (b) Night crew in well A; (c) Day crew in well B; (d) Night crew in well B; (e) Day crew in well C; and (f) Night crew in well C.*

6.2. ILT Measurement

In order to compute ILT, an optimal KPI for connection time needs to be fixed. This KPI serves as a reference or benchmark for real-time monitoring [26, 33]. Various evaluation methods, such as the best composite, average, and P50 can be used to determine the target KPI [34].

In this paper, the P50 distribution and the best composite were applied to define the KPIs for connection time. Individual histograms for "weight to slip" KPI, "slip to slip" KPI, and "slip to weight" KPI are depicted in **Fig. 6.** From this **Fig. 6**, it can be observed that the P50 is 21.33 minutes for "weight to slip" time, 7.42 minutes for "slip to slip" time, and 4 minutes for "slip to weight" time. As a result, the KPI for "weight to weight" is determined to be 32.75 mins.

Following the selection of the KPI, ILT and the potential saving time were determined using the defined weight to slip KPI, slip to slip KPI, and slip to weight KPI. ILT is expressed here as the cumulative difference between the actual operation time and the target KPI. **Table 3** highlights the analysis results of ILT during connection operations.

Table 4 and **Table 5** indicate the total saving if all the connections in the three wells were carried out at the KPIs defined in this study. Applying the P50 distribution, the total saving time would be 2 days, 16 hours, and 57 minutes, representing 29.96% of the total time spent in the three wells for the connection operation. Applying the best composite, the total saving time would be 6 days, 4 hours, and 53 minutes, representing 68.67% of the total time spent in the three wells for the connection operation.

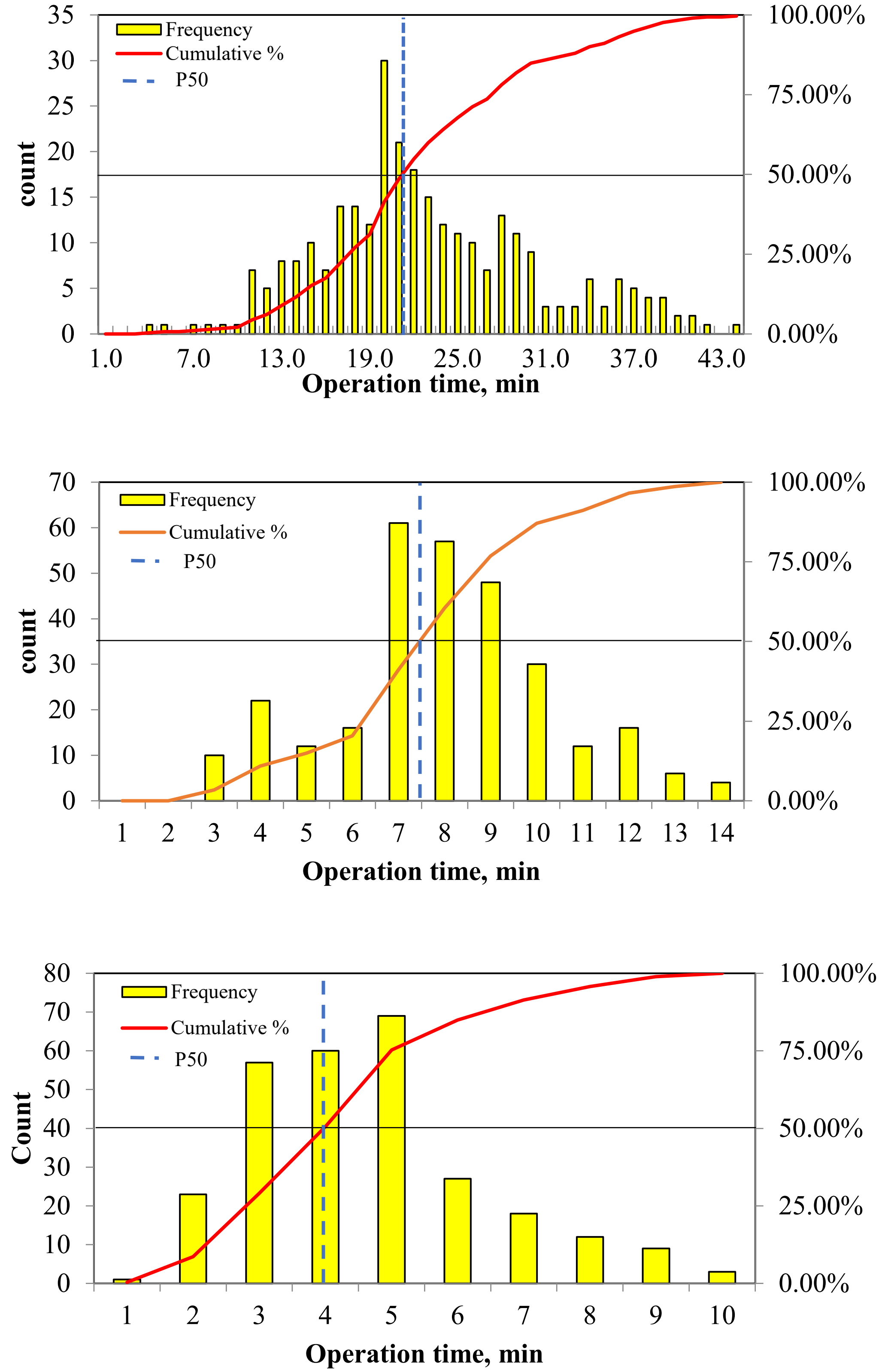


***Fig. 6** Determination of the KPIs for connection time : (a) weight to slip KPI ; (b) Slip to slip KPI; (c) Slip to weight KPI.*

Table 4 Saving time for weight to weight connection in the three wells using the P50 distribution.

	slip to weight	slip to slip	slip to weight	weight to weight
Operation time	5 days, 12 hrs, 39 mins	2 days, 3 hrs, 53 mins	1 day, 8 hrs, 16 mins	9 days, 49 min
KPI	7.58 mins	2.41 mins	1 mins	8 mins
Savings potential time	3 days, 21 hrs, 21 mins	1 day, 7 hrs, 31 mins	1 day	6 days, 4 hrs, 53 mins
Savings potential time (%)	70.38%	60.75%	74.36%	68.67%

Table 5 Saving time for weight to weight connection in the three wells using the best composite.

	slip to weight	slip to slip	slip to weight	weight to weight
Operation time	5 day, 12 hrs, 39 mins	2 day, 3 hrs, 53 mins	1 day, 8 hrs, 16 mins	9 days, 49 min
KPI	21.33 mins	7.42 mins	4 mins	34 mins
Savings potential time	1 day, 9 hrs, 7 mins	17 hrs, 34 mins	14 hrs, 16 mins	2 day, 16 hrs, 57 mins
Savings potential time (%)	24.97%	33.87%	44.21%	29.96%

7. Conclusions

The present work applied the Naive Bayes algorithm to create a novel robust rig state, leveraging training data generated from drilling experts' recommendations. The developed model demonstrated an impressive 99.747% success rate in accurately identifying drilling operations. The following are the main findings that can be drawn from this study:

1. Instead of relying on surface data for training rig state models, the study emphasized the significance of employing data that contains insights into the cognitive processes of drilling engineers for recognizing drilling operations using surface data.

2. The Naive Bayes algorithm, known for its simplicity and ease of implementation, demonstrated excellent performance in recognizing the rig state compared to other complex classification approaches.
3. The established rig state successfully processed high-frequency time series data collected from surface sensors in three wells. This enabled the analysis of drilling crew performance and the derivation of KPIs related to efficiency.
4. The finding of this study highlights the effectiveness of the established rig state model in supporting drilling supervisors in achieving efficiency gains and making decisions. Integrating this model in Real-Time Operation Monitoring (RTOM) projects holds promise for continual improvements in operational performance, contributing to time and cost savings.
5. The proposed rig state model is not applicable for directional drilling since it does not consider some rig states encountered in drilling with mud motors. Further work should address this limitation by updating the training data and extending its application.
6. Overall, this research has demonstrated the potential of ML algorithms in monitoring drilling operation and enhancing its efficiency. Future investigations should explore additional data-driven models and diverse data sources to further improve the accuracy and robustness of rig state models.

Abbreviations

ANN	Artificial Neural Network
Acc	Accuracy
CNN	Convolutional Neural Network
DT	Decision Tree
KPI	Key Performance Indicator
LSTM	Long-Short-Term Memory
ILT	Invisible Lost Time
ML	Machine Learning
NB	Naive Bayes
NPT	Non-Productive Time
POOH	Pulling Of Out Hole
RIH	Running In Hole
RNN	Recurrent Neural Network
RPM	Revolutions Per Minute
RTOM	Real Time Operation Monitoring
WOH	Weight On Hook

Highlights

- A novel robust rig state was developed using the Naive Bayes classifier.
- Trustworthy data that provides sufficient insights into rig state recognition was used in this paper to train the rig state classifier.
- The developed rig state model was applied to label comprehensive field data and extract valuable insights, such as the KPIs and ILT.

- The findings of this study demonstrate the vital role of integrating the developed rig state model into Real-Time Operation Monitoring projects.

Appendix A. Theory: rig state

The rig state classifier is a system that uses the data recorded by the surface sensors to automatically detect the drilling operation performed by the drilling crew. By setting the KPI as the maximum time to be respected throughout the execution of each operation, the real-time automatic detection of the rig state has the potential to quantify the invisible lost time. Furthermore, the rig state classifier facilitates the data segmentation for any drilling data processing task. For example, a well drilled for over 60 days with a sample rate of 1 Hz will generate big data (nearly 5 million data points). To segment this data, the user must manually filter the data, which is both complex and time-consuming. Using an automatic drilling operations classifier, this process becomes fairly simple and may be completed in a matter of seconds. In this paper, the Naïve Bayes algorithm was employed to create a trustworthy classifier that can accurately detect and recognize the drilling operation based on surface data.

Appendix B. Theory: Naive Bayes Classifier

The Naive Bayes method relies on the Bayes theorem with the assumption of strong independence between variables [35, 36]. The latter assumes that the existence of one object's feature (attribute) in a class is independent of the presence of another feature, which means that all the features contribute independently to the classification of an object even if they are dependent on each other [37, 38], and this is why this approach is referred to as naive.

Assuming a classification problem in which X=$X1$,$X2$,⋯,Xn represents an observation with n independent attributes, and C_j is one of K classes. Using Bayes' theorem, the Naive Bayes Classifier computes conditional probabilities $P(Cj/X1, X2, , Xn)$ for each class. From previous knowledge, Bayes' theorem allows us to determine the posterior probability of the class C_j knowing the attribute X of an object or individual as follows [31, 39]:

$$P(C_j/X) = \frac{P(C_j)P(X/C_j)}{P(X)} \tag{B.1}$$

where $P(C_j)$ stands for the priori probability of Cj, $P(X)$ is the prior probability of X or the marginal probability, $P(X/C_j)$ is the likelihood of observing X knowing that the class is C_j.

Applying the assumption of the independence of variables X=$X1$,$X2$,⋯,Xn, the likelihood $P(X/C_j)$ can be computed using the following expression :

$$P(X/C_j) = \prod_{i=1}^{n} P(X_i/C_j) \tag{B.2}$$

by substituting the posterior probability of the class C_j with its expression, **Eq.(B.1)** can be expressed as follows :

$$P(C_j/X) = \frac{P(C_j)\prod_{i=1}^{n} P(X_i/C_j)}{P(X)} \tag{B.3}$$

Following the computation of conditional probabilities $P(C_j/X)$ for K classes, the observation X will be attributed to the class C if the following condition is met:

$$P(C/X) \geq P(C_j/X) \text{ for each } 0 \leq j \leq 1 \tag{B.4}$$

A Naive Bayes model is trained by computing and storing a set of probabilities based on the training data. This list includes:

- The prior probabilities $P(C_j)$: are simply the number of observations belonging to the class C_j divided by the total number of observations m.

$$P(C_j) = \frac{\sum_{i=1}^{m} I(y_i = C_j)}{m} \tag{B.5}$$

- The likelihoods $P(X_i/C_j)$: Given that each feature X can take on a large number of possible values, the likelihood ratio of C_j for a given feature value a_d is calculated by dividing the frequency of observations in which that feature value belongs to the class C_j by the frequency of all observations belonging to the same class C_j.

$$P(X_1 = a_d/C_j) = \frac{\sum_{i=1}^{m} I(X_{1i} = a_d, y_i = C_j)}{\sum_{i=1}^{m} I(y_i = C_j)} \tag{B.6}$$

After the training phase, the developed Naive Bayes model predicts the class y of the observation X as follows [40]:

$$y = C \qquad \text{if} \qquad \mathrm{P}(C)\prod_{i=1}^{n} P(X_i/C) = max\left(P(C_j)\prod_{i=1}^{n} P(X_i/C_j)\right) \tag{B.7}$$

Declarations

- **Competing Interests**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

- **Authors contribution statement**

All authors contributed to the study conception and design. Data collection were performed by Farouk Said Boukredera. Formal analysis and investigation were conducted by Mohamed Riad Youcefi, Farouk Said Boukredera, Khaled Ghalem and Chinedu Pascal Ezenkwu under the supervision of Ahmed Hadjadj. The first draft of the manuscript was written by Mohamed Riad Youcefi and was edited and reviewed by Chinedu Pascal Ezenkwu. All authors read and approved the final manuscript.

- **Ethical and informed consent for data used**

This research does not contain any studies that involve human and/or animal participants, their data, or biological material.

- **Data availability and access**

The datasets analyzed during the current study are not publicly available due to the restrictions applied to the availability of these data, which were used under license for the current study, and so are not publicly available.

- **Funding statement**

No funding was received for conducting this study.

References

1. Crow DJG, Anderson K, Hawkes AD, Brandon N (2018) Impact of Drilling Costs on the US Gas Industry: Prospects for Automation. Energies 11:2241
2. Raimi D, Krupnick AJ, Shah J-S, Thompson A (2021) Decommissioning orphaned and abandoned oil and gas wells: New estimates and cost drivers. Environ Sci Technol 55:10224–10230
3. Barakat M, Abu El Ela M, Khalaf F (2021) Integrating risk management concepts into the drilling non-productive time. J Pet Explor Prod 11:887–900
4. Ashena R, Elmgerbi A, Rasouli V, et al (2020) Severe wellbore instability in a complex lithology formation necessitating casing while drilling and continuous circulation system. J Pet Explor Prod Technol 10:1511–1532
5. Magzoub MI, Salehi S, Hussein IA, Nasser MS (2020) Loss circulation in drilling and well construction: The significance of applications of crosslinked polymers in wellbore strengthening: A review. J Pet Sci Eng 185:106653
6. Yang X, Zhang Y, Zhou D, et al (2023) Drilling conditions classification based on improved stacking ensemble learning. Energies 16:5747
7. Forshaw M (2022) Case Study: Automated Stringer Detection Boosts Drilling Efficiency With Less Downtime. J Pet Technol 74:39–43
8. de Oliveira GL, Zank CA, Costa AF, et al (2016) Offshore Drilling Improvement Through Automating the Rig State Detection Process-Implementation Process History and Proven Success. In: IADC/SPE Drilling Conference and Exhibition. OnePetro
9. Sarker IH (2021) Machine learning: Algorithms, real-world applications and research directions. SN Comput Sci 2:1–21

10. Bertolini M, Mezzogori D, Neroni M, Zammori F (2021) Machine Learning for industrial applications: A comprehensive literature review. Expert Syst Appl 175:114820

11. Jain R, Gupta M, Taneja S, Hemanth DJ (2021) Deep learning based detection and analysis of COVID-19 on chest X-ray images. Appl Intell 51:1690–1700

12. Zhang W, Gu X, Hong L, et al (2023) Comprehensive review of machine learning in geotechnical reliability analysis: Algorithms, applications and further challenges. Appl Soft Comput 110066

13. Hegde C, Millwater H, Gray K (2019) Classification of drilling stick slip severity using machine learning. J Pet Sci Eng 179:1023–1036

14. Hegde C, Millwater H, Pyrcz M, et al (2019) Rate of penetration (ROP) optimization in drilling with vibration control. J Nat Gas Sci Eng 67:71–81

15. Hegde C, Pyrcz M, Millwater H, et al (2020) Fully coupled end-to-end drilling optimization model using machine learning. J Pet Sci Eng 186:106681

16. Boukredera FS, Youcefi MR, Hadjadj A, et al (2023) Enhancing the drilling efficiency through the application of machine learning and optimization algorithm. Eng Appl Artif Intell 126:107035

17. Muojeke S, Venkatesan R, Khan F (2020) Supervised data-driven approach to early kick detection during drilling operation. J Pet Sci Eng 192:107324

18. Gurina E, Klyuchnikov N, Antipova K, Koroteev D (2022) Forecasting the abnormal events at well drilling with machine learning. Appl Intell 1–16

19. Tran T, Pham NT, Lundgren J (2022) A deep learning approach for detecting drill bit failures from a small sound dataset. Sci Rep 12:9623

20. Mopuri KR, Bilen H, Tsuchihashi N, et al (2022) Early sign detection for the stuck pipe scenarios using unsupervised deep learning. J Pet Sci Eng 208:109489

21. Lakhanpal V, Samuel R (2017) A New Approach to Harness Data for Measuring Invisible Lost Time in Drilling Operations. In: SPE Annual Technical Conference and Exhibition. OnePetro

22. Al Ady A, Franco N, Corona M, Dorantes A (2023) Operation Efficiency and Rig Performance Improvements through Data Analytics. In: SPE Gas & Oil Technology Showcase and Conference. SPE, p D011S010R006

23. Zhao J, Chambon S, Shen Y, et al (2019) Automatic Slip Status and Stand Detection in Real-Time Drilling. In: Offshore Technology Conference. OnePetro

24. Coley C (2019) Building a Rig State Classifier Using Supervised Machine Learning to Support Invisible Lost Time Analysis. In: SPE/IADC International Drilling Conference and Exhibition. Society of Petroleum Engineers

25. Ben Y, James C, Cao D (2019) Development and application of a real-time drilling state classification algorithm with machine learning. In: SPE/AAPG/SEG Unconventional Resources Technology Conference. OnePetro

26. Yin Q, Yang J, Hou X, et al (2020) Drilling performance improvement in offshore batch

wells based on rig state classification using machine learning. J Pet Sci Eng 192:107306

27. Arnaout A, Fruhwirth R, Esmael B, Thonhauser G (2012) Intelligent real-time drilling operations classification using trend analysis of drilling rig sensors data. In: SPE Kuwait International Petroleum Conference and Exhibition. OnePetro

28. Nhu V-H, Shirzadi A, Shahabi H, et al (2020) Shallow landslide susceptibility mapping: A comparison between logistic model tree, logistic regression, naïve bayes tree, artificial neural network, and support vector machine algorithms. Int J Environ Res Public Health 17:2749

29. Bustamam A, Bachtiar A, Sarwinda D (2019) Selecting features subsets based on support vector machine-recursive features elimination and one dimensional-naïve bayes classifier using support vector machines for classification of prostate and breast cancer. Procedia Comput Sci 157:450–458

30. Bogdal C, Schellenberg R, Höpli O, et al (2022) Recognition of gasoline in fire debris using machine learning: Part I, application of random forest, gradient boosting, support vector machine, and naïve bayes. Forensic Sci Int 331:111146

31. Phoenix P, Sudaryono R, Suhartono D (2021) Classifying promotion images using optical character recognition and Naïve Bayes classifier. Procedia Comput Sci 179:498–506

32. Akinfolarin A, Amah O, Ijeh I, et al (2020) Lessons Learned from Drilling a Long Open Hole Interval and Recovery from a Stuck Pipe Incident. In: SPE Nigeria Annual International Conference and Exhibition. SPE, p D013S002R002

33. Haneef J, Sheraz A (2023) A Comparative Analysis of Well Key Performance Indicators (KPIs) with Well Complexities Using Well Complexity Calculator. Arab J Sci Eng 48:9339–9356

34. Al-Ghunaim SM, Sadiq BM, Siam M, et al (2017) Operations Efficiency: Improved Well Planning Methodology Based on Invisible Lost Time Smart KPIs. In: SPE Middle East Oil & Gas Show and Conference. OnePetro

35. Swinburne R (2004) Bayes' Theorem. Rev Philos la Fr l 194:

36. Joyce J (2003) Bayes' theorem

37. Blanquero R, Carrizosa E, Ramírez-Cobo P, Sillero-Denamiel MR (2021) Variable selection for Naïve Bayes classification. Comput Oper Res 135:105456

38. Kwon Y, Kwasinski A, Kwasinski A (2019) Solar irradiance forecast using naïve Bayes classifier based on publicly available weather forecasting variables. Energies 12:1529

39. Maswadi K, Ghani NA, Hamid S, Rasheed MB (2021) Human activity classification using Decision Tree and Naive Bayes classifiers. Multimed Tools Appl 80:21709–21726

40. Shang F, Li Y, Deng X, He D (2018) Android malware detection method based on naive Bayes and permission correlation algorithm. Cluster Comput 21:955–966