République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Amar Telidji- Laghouat

FACULTE : TECHNOLOGIE

DEPARTEMENT : GÉNIE DES PROCÉDÉS

MEMOIRE DE MASTER

Présenté par : SELLAM kholoud

DOMAINE : Sciences et Technologies

FILIERE : Hydrocarbures

OPTION : Génier Gazier

Thème

Development of an intelligent system for measuring invisible lost time during drilling operations

Jury de soutenance :

Nom et Prénom	Grade	Qualité
HADJADJ Asma	MCB	Président
GHALEM Khaled	MAA	Examinateur
MERIGUI Khaled	MAA	Examinateur
YOUCEFI Mohamed Riad	MCB	Rapporteur

Année Universitare : 2022-2023

عنوان المذكرة: تطوير نظام ذكي لقياس الوقت الضائع غير المرئي أثناء عمليات الحفر.

اللقب: سلام **الإسم:** خلود **المؤطر:** يوسفي محمد رياض

ملخص يعد تحسين عملية الحفر ومراقبتها خطوة حاسمة في صناعة النفط والغاز. أحد العوامل التي قد تسبب ارتفاعًا هائلاً في تكلفة الحفر هو الوقت الضائع غير المرئي (ILT). تهدف هذه الأطروحة إلى تطوير نظام ذكي قادر على قياس ILT تلقائيًا في الوقت الفعلي. لهذا الغرض ، أنشأنا نموذج تصنيف الحفر باستخدام احد خوارزميات الذكاء الاصطناعي (SVM). تم جمع البيانات من بئرين لتقييم مؤشر الأداء الرئيسي خلال عملية إضافة أعمدة الحفر. أخيرًا ، تم قياس الوقت الضائع غير المرئي أثناء اتصال عملية في البئر أ و ب. وأظهرت النتائج أن إجمالي الوقت الضائع غير المرئي هو ساعتان و 53 دقيقة ، أي ما يعادل 22.67٪ من إجمالي الوقت الذي يقضيه في إضافة أعمدة الحفر. بالنسبة للبئر B ، يبلغ إجمالي ILT ساعة و 44 دقيقة ، وهو ما يعادل 19.35٪ من إجمالي الوقت الذي يقضيه في إضافة أعمدة الحفر. يمكن أن يساعد اكتشاف هذه الأطروحة بشكل كبير شركات الحفر على تحسين عملياتها وتقليل التكاليف وتحسين الأداء العام.

كلمات مفتاحية : الحفر الدوار, التعلم الآلي , الوقت الضائع غير المرئي ,نظام ذكي , عمليات الحفر.

Memory title : Development of an intelligent system for measuring invisible lost time during drilling operations.

Name: Sellam **First name:** Kholoud **Directed by:** Mohamed Riad Youcefi

Abstract : Optimizing and monitoring the drilling process is a crucial step in the oil and gas industry. One of the factors that might cause an immense rise in the cost of drilling is Invisible Lost Time (ILT). This thesis aims to develop an intelligent system capable of automatically measuring the ILT in real time. For that purpose, we created a drilling classification model using Support Vector Machine (SVM). The, data from two wells were gathered to assess the drilling connection's Key Performance Indicator (KPI). Lastly, the ILT during drilling connection were measured in well-A and well-B. The results showed that the total ILT is 2 hours and 53 minutes, which is equivalent to 22.67 % of the overall time spent for making drilling connections. For well-B, the total ILT is 1 hours and 44 minutes, which is equivalent to 19.35 % of the overall time spent for making drilling connections. The finding of this thesis can significantly helps the drilling companies to optimize their operations, reduce costs, and improve overall performance.

Key words: Drilling, Invisible Lost Time, Artificial Intelligence, Machine Learning, SVM, Rig State.

Titre du mémoire : Développement d’un système intelligent de mesure du temps perdu invisible lors des opérations de forage.

Nom : SELLAM **Prénom :** Kholoud **Encadreur :** YOUCEFI Mohmed Riad

Résumé L'optimisation et le suivi du processus de forage est une étape cruciale dans l'industrie pétrolière et gazière. L'un des facteurs susceptibles d'entraîner une augmentation considérable du coût du forage est le temps perdu invisible (ILT). Cette thèse vise à développer un système intelligent capable de mesurer automatiquement l'ILT en temps réel. Pour ce faire, nous avons créé un modèle de classification des forages en utilisant le Support Vector Machine (SVM). Ensuite, les données de deux puits ont été collectées pour évaluer les indicateurs clés de performance (KPI) de l’opération de l’ajout des tiges. Enfin, les ILT pendant l’ajout des tiges ont été mesurés dans les puits A et B. Les résultats ont montré que l'ILT total est de 2 heures et 53 minutes, ce qui équivaut à 22,67 % du temps total passé à effectuer pour l’ajout des tiges. Pour le puits B, l'ILT total est de 1 heure et 44 minutes, ce qui équivaut à 19,35 % du temps total consacré à pour l’ajout des tiges. Les résultats de ce mémoire peuvent aider de manière significative les entreprises de forage à optimiser leurs opérations, à réduire les coûts et à améliorer les performances globales.

Mots clés : Forage, ILT, Intelligence Artificielle, Apprentissage automatique, SVM, logiciel de la détection de l’état de l’appareil de forage.

Table of Contents

List of Figures

List of Tables

Tableau IV.1. Training data.

Tableau IV.2. ILT for well-A and well-B.

List of Abbreviations

AI: Artificial Intelligence.

ML: Machine Learning.

NPT: Non Productive Time.

ANN: Artificial Neural Network.

ILT: Invisible Lost Time.

KPI: Key Performance Indicators.

CNN: Convolutional Neural Network.

RNN: Recurrent Neural Network.

WOB: Weight on Bit.

ROP: Rate of Penetration.

WOH 1: stands for the weight on hook when the drill string when it is suspended.

WOH 2: represents the weight on hook during drilling.

ROM: Revolution per Minute.

WBM: Water Based Mud.

OBM: Oil Based Mud.

SBM: Synthetic Based Mud.

POOH: Pull out of the hole.

RIH: Run in the Hole.

SVM: Support Vector Machines.

Acknowledgments

“First, we would like to thank **ALLAH**, our Creator to give us the chance to study and strength to do this study”.

“Our sincere thanks to My **Father**, **Mother**, Baba, and My Sister and two Brothers”.

“Our sincere thanks to our Professor Dr. Mohamed Riad Youcefi”.

“Our last thanks are not the least, go to all those who have contributed from near or far for the completion of this work”.

Introduction General

Oil and gas drilling operations are intricate and time-sensitive, necessitating careful planning and execution. However, a number of variables may play a role in the waste of crucial time during drilling operations, which will result in lower productivity and higher expenses. Optimizing drilling processes requires minimizing invisible wasted time, which refers to inefficiencies that are not immediately evident or simple to observe. Machine learning algorithms have become effective tools for data analysis and the extraction of insightful information in recent years. Utilizing these methods, especially Python programming, can assist in locating and resolving the underlying causes of lost time in drilling operations.

This thesis aims for identifying and addressing the hidden sources of lost time, so operators can make informed decisions to optimize drilling processes, enhance overall efficiency, and reduce operational costs.

The ultimate aim of this project is to empower drilling operators to make data-driven decisions to reduce inefficiencies by giving them important insights into the underlying causes of lost time. Drilling procedures may be optimized using Python's machine learning capabilities, resulting in higher output, lower costs, and better performance overall.

Machine learning techniques, such as SVM, provide a data-driven approach to analyzing drilling data and predicting potential sources of lost time. SVM is a supervised learning algorithm that can effectively classify data into different categories based on patterns and features. By training an SVM model with historical drilling data that includes both successful and time-consuming operations, the algorithm can learn to recognize patterns indicative of potential lost time scenarios.

Chapter 1 : Literature review - AI applications for ILT minimization

1. AI applications in drilling industry

In recent years, drilling has developed considerably. With drilling activities now generating more data and transmitting multiple observations in real time (Sadanandan, 2014), it has become necessary to understand and exploit this Big Data by adapting the fourth industrial revolution to minimize the cost of drilling. The use of advanced Artificial Intelligence (AI) approaches such as Machine Learning (ML) models and various optimization strategies are key to improving the efficiency of real-time drilling operations. AI involves the use of advanced methods and algorithms capable of solving multi-dimensional challenges analogous to human cognitive abilities.

In recent years, algorithms such as artificial neural networks, fuzzy logic and genetic algorithms have become extremely popular in the oil industry, particularly in drilling to mitigate real-time drilling challenges and achieve higher penetration rates and lower Non Productive Time (NPT). According to[2], machine learning aims to exploit the large amounts of data generated every second by the drilling data acquisition system to reduce the risks associated with drilling operations. AI was first applied in the oil industry in 1898, where artificial neural networks were used to interpret drilling logs, diagnose tools, monitor important drilling and completion operations and simulate reservoirs (Bello et al., 2016).

This section of the thesis is conacred to discuss various advanced applications and artificial intelligence approaches that was applied to minimize the invisible lost time in real time.

2. Building a Rig State Classifier Using Supervised Machine Learning to Support Invisible Lost Time Analysis.

(Coley, 2019) Reported in their study that the construction of an internal system to assess the invisible lost time is essential for minimizing the NPT and the drilling costs. They proposed in their study the creation of based on rig state based on the use of supervised machine learning. The authors affirmed that identifying rig-state is at the core of every performance and engineering analysis system. The project's goal was to provide efficiency and engineering metrics in a central repository that covers both historic and near real-time analysis to deliver a rich resource for offset comparison.

The output of this study indicated that building a rig-state classifier's can concurrently support all real-time activities, analyse historical well data quickly for offset benchmarking, cloud storage and processing, near-real-time and historical analysis capabilities, cheap cost, high accuracy, and consistent results are also possible.

3. Development and Application of a Real-Time Drilling State Classification Algorithm with Machine Learning

Another study was proposed by (Ben et al., 2019) to confirm the importance of an automatic rig state detection for real-time drilling analytics system. The authors applied several machine learning algorithms, including Random Forest, Convolutional Neural Network (CNN), and a hybrid Convolutional Neural Network/Recurrent Neural Network (CNN/RNN) to build drilling operation classification models.

The data used in this study were collected from two field, a total of 22 wells with 11 million rows of data. It was discovered that learning models were considerably superior to rules-based models. Over 99% of the models used machine learning.

After developing a reliable rig state model, this letter was used to evaluate High-frequency time-series data with various rig states such as slide drilling, rotate drilling, reaming, slack off, pick up, and in slips. This labelled time-series data was used to obtain drilling and tripping connection parameters. Driller and rig supervisor actions have resulted in up to 70% savings in connection times due to the introduction of the drilling and tripping connection criteria. The output of the rig-state classifier was then used to create KPI data to facilitate ILT analysis.

4. Drilling performance improvement in offshore batch wells based on rig state classification using machine learning.

(Yin et al., 2020) applied the Artificial Neural Network (ANN) model for classifying rig states, for assessing the rig crew performance. The established ANN model was applied then to examine the operating time, display histograms for the operational time, and asses the performance of the rig crew. The Invisible Lost Time (ILT) is also identified and minimized in this study by comparing the operating time and Key Performance Indicators (KPIs) (as developed and established by the operator). The authors affirmed in their study that the created ANN model achieved an accuracy rate of 93%. In addition, the authors stated that by using the ANN model for evaluating the performance of the rig crew, the ILT was reduced by 45.23%, and total drilling performance was improved by 31.19%.

5. Contribution and novelty

The majority of works presented in the literature presented a classification model with significant errors. To our knowledge, this is due to the fact that they used drilling data that was recorded by the rig sensors to train their classification models. Even if the calculation errors are minor, they might lead to an inaccurate assessment of the KPI and ILT when the built system is used to assess the drilling efficiency.

In this thesis, a novel approach for improving rig state performance is provided. The Support Vector Machine (SVM) is used to create an intelligent system capable of recognizing real-time drilling activities and assessing the ILT. This study differs from earlier ones in that the created rig state is trained using data based rules rather than drilling data, which is recommended here to improve precision when classifying drilling operations.

The results of this thesis are destined to prove the potential of AI methods to automate the process of monitoring drilling operations, minimize the NPT, and reduce drilling costs.

Chapter 2 : Drilling - Equipment and Operations

1. Introduction

Rotary drilling remains the essential process for exploring and exploiting oil and gas fields, enabling us to reach great depths under optimum technical and safety conditions. This technique involves applying force and rotation to a drill tool mounted on the bottom end of a drill string. The combination of weight on bit (WOB) and the rotation allows the rock destruction. Rotation is transmitted from the top drive or rotation table at the surface to the drilling tool via the drill string. The weight is provided by a set of drill collars which form the lower part of the drill string. Drill pipes are also used to circulate high-pressure drilling fluid (mud). This mud cools the drill bit and ensures that rock debris is evacuated. Casings are run into the well and cemented in place to ensure the well's long-term stability and protect it against formation fluid invasion.

2. Drilling equipment

A drilling system consists of two parts: the drilling rig and the drill string:

- The drilling rig consists of a tower, known as a derrick , on which the lifting equipment for handling, screwing and unscrewing the pipes, and changing the drill bit are positione.
- The component that functions inside the well is know as the drill string, it serves primarily as:
    - Provide the energy required for rock destruction.
    - Control the well trajectory.
    - Transmit the hydraulic force.
    - Transmit the rotation movement from the surface to the drill bit

3. Drilling functions

The following activities are carried out during the rotational drilling of a deep well :

- Run the drill pipe in the well during the drilling and connection operation.
- Pull out the drill pipe out of the well.
- Ensure the drill pipe rotation.

- Ensure the injection of the drilling fluid into the well to elevate the cut cuttings, cool the drill bit, and support the wellbore.
- Run the casing in the hole (CHERIFI, 2012).

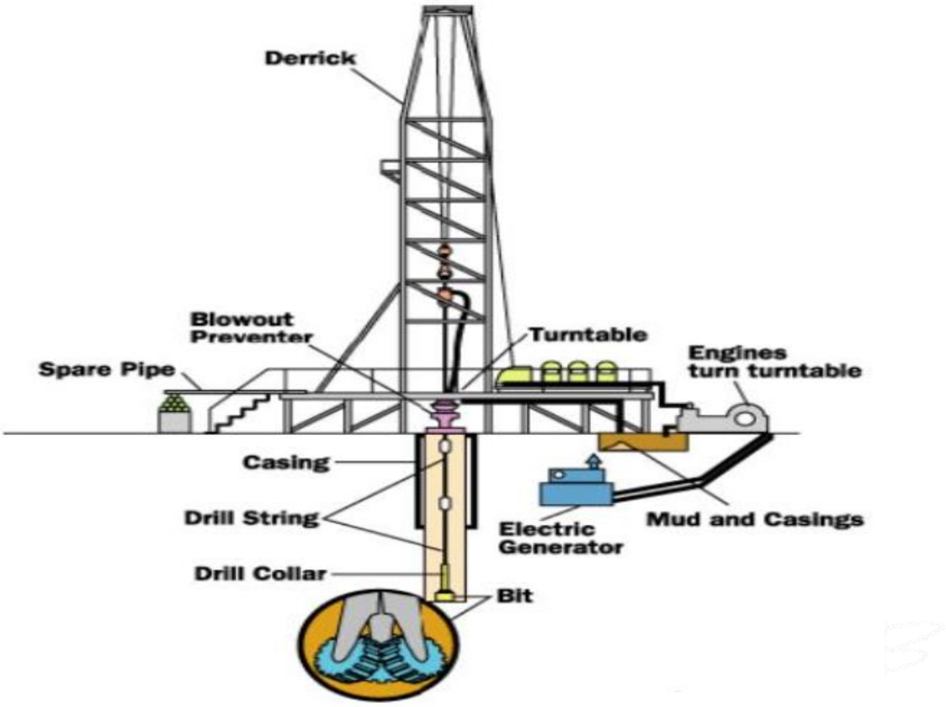

Figure I.1. Drilling rig & drill string composition

All of this above-ground equipment serves three primary tasks:

3.1. Rotating function

To rotate the drill bit, a cylindrical, square or hexagonal-sectioned pipe, called the Kelly, is connected at the top of the drill string, and into a surface equipment known as the rotary table(CHERIFI, 2012).

3.2. Lifting function:

To lifting the drill string (drill pipe – drill collars – heavy weight pipe), a high capacity crane must be used, since the drilling string might weigh up to 150 tonnes. This crane consists of (CHERIFI, 2012):

- The derrick
- The drawworks
- The traveling block
- Crown block
- The cable

The drawworks is typically powered by an electric or diesel engine. The drawworks consists of a large drum around which the drilling cable is spooled. By rotating the drum, the drawworks creates the necessary tension in the drilling line, allowing it to lift and lower loads.

The crown block is placed near the top of the drilling rig's derrick. It consists of several sheaves or pulleys through which the drilling cable passes. Its main purpose is to change the direction of the drilling cable from vertical to horizontal, facilitating the lifting operation.

The traveling block is connected to the crown block via the drilling cable. It moves vertically along the derrick for carrying the hook, which is attached to the load being lifted.

Figure I.2. Lifting function equipment's : (a) crown block; (b) traveling block; (c) drawworks

4. The drilling parameters:

The drilling parameters can be divided into two classes : the mechanical and hydraulic parameters. The most well known mechanical parameters is the Rate of Penetration (ROP). ROP optimization is one of the most essential aspects of the drilling economics since drilling time accounts for around half of the entire cost of the well. Increased ROP increases drilling efficiency and minimize the costs dramatically.

4.1. Mechanical parameters

The value of these mechanical characteristics varies depending on the hardness of the geological strata penetrated by the drill bit.

4.1.1. Weight On Bit

The weight on the (WOB) tool reflects the weight applied to the bit, as computed using the method below. The required WOB increases with increasing the rock compressibility resistance.

$$\mathbf{WOB = WOH\ 1 - WOH\ 2}$$

where WOH 1 stands for the weight on hook when the drill string when it is suspended, and WOH 2 represents the weight on hook during drilling (Vempati et al., 2020).

4.1.2. Revolution per Minute (RPM)

The rotation speed or RPM denotes the number of turns per minute of the rotation table on the rig floor, which ranges from 50 to 300 spins per minute depending on the hardness of the formation.

4.1.3. Torque

Torque is not a characteristic to optimize; it is simply the opposing force of rotation, and it is primarily determined by the weight on the tool (WOB), rotation speed (RPM), and hardness of the formation. Surface torque is not communicated by the drilling tool, yet surface measurement is now the only option (Vempati et al., 2020).

4.2. Hydraulic parameters

Hydraulic parameters are all factors related to the drilling fluid such as flow rate, pressure, drilling fluid type, and drilling fluid characteristics (density, viscosity, etc.) that play the roles of lubrication, cooling, and bottom hole cleaning and thus affect the speed of advancement and tool lifetime.

4.2.1. Type de drilling mud

The type of drilling mud is chosen according to the desired performance and designates the physical-chemical properties of the drilling fluid. Three types of sludge are often Water Based Mud (WBM), Oil Based Mud (OBM) and Synthetic Based Mud (SBM). A mud is a mixture of water and chemical additives.

4.2.2. Flow and hydraulic pressure

The flow rate and the hydrostatic pressure are the physical variables that must properly chosen to ensure an appropriate cuttings removal while avoiding undesirable events like the kick, formation collapse, and the stuck pipe.

4.2.3. Mud density

The density of the drilling mud provides information about the well, notably the regulation of the pressure in the well. In addition to the evacuation of the cutting from the bottom hole to the surface, the drilling mud is designed for maintaining the reservoir fluids such as the shallow gas stored in the rocks [8].

5. The principals drilling operations:

5.1. Drilling

The time spent drilling through formations to reach the target formation is referred to as the drilling operation. The rotary table or top drive turns at this point, causing the bit to rotate through the drill string. In addition to the rotating function, the driller controls and limits the hook's descent by pressing the brake. During this procedure, the driller maintains a set WOB depending on the compression resistance of the rock. The driller observes the weight indicator (Martin Decker) and must maintain it constantly by leaving the Kelly descent at the same pace as the drill bit penetration. The RPM and flow rate are two more characteristics that must be managed during drilling. These two parameters are normally set, and the driller checks and modifies their values in accordance with the program, while also ensuring that the flow pressure remains within acceptable limits .

5.2. Drilling connection

When the drill bit has drilled a length of drill pipe of 30 ft, the drill string must be extended by the same amount by inserting a new drill pipes under the kelly. The various sequences are described in Figure I.3.

During drilling, the workers place a drill pipes in a mouse-hole located close to the rotation table. The foreman engages the drawworks to lift the drill string and place to the first drill pipe below the rotation table. The workers put the slips in place, and the kelly can be unscrewed, as the drill string is now suspended on the rotary table. Naturally, mud circulation is then stopped. In Figure I.3b, the workers connect the kelly to the drill pipe inserted in the mouse-hole, than the foreman lifts the kelly and drill pipe assembly (Figure I.3c).

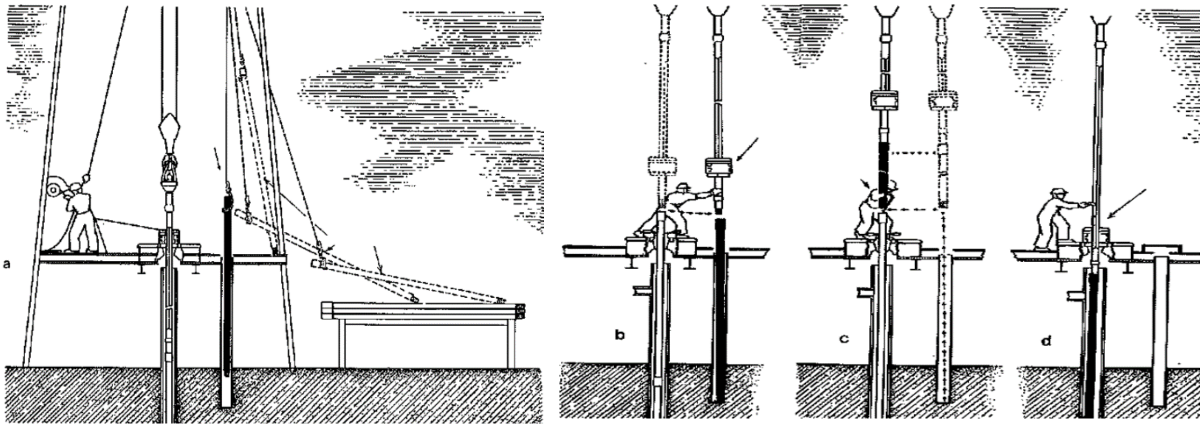

Figure I.3. Procedure of the drilling connection (Nguyen, 1993)

Once the new drill pipes has been connected to the drill string, the foreman restarts the circulation of drilling fluid. The driller positions the kelly in the rotary table and drilling can resume (Figure I.3d).

5.3. Tripping operation

When the tool has become wearable, or when the desired depth has been reached, the entire assembly must be lifted up, either to change the drill bit or to run in the casing. This operation is know as Pull out of the hole (POOH) (Nguyen, 1993).

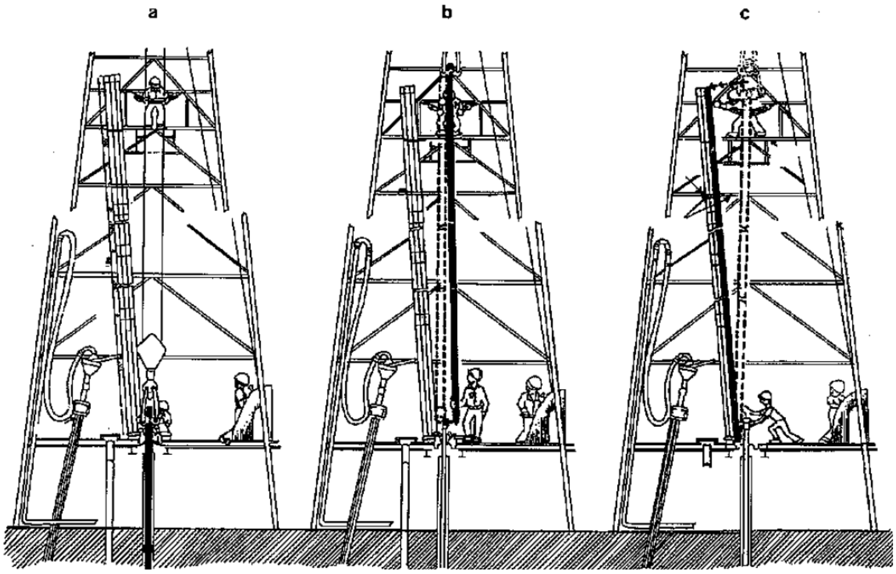

Figure I.4. Procedure of the tripping operation (Nguyen, 1993)

The first operation is to remove the injection head from the drilling hook and store the kelly and injection head assembly, still connected to the pumps, in a sheath called a rat-hole (Figure I.4a). The workers close the elevator under the tool-joint of the first drill pipe, and the foreman operates the drawworks to raise the drill string to a height corresponding to three drill pipes. The fourth pipe is fixed in the table by the slips, and this connection is unscrewed with keys (Figure I.4b). A set of three pipes is then suspended from the elevator. The workers push back the lower end of this length (stand) to support it on a stacking frame (set back). As soon as this is done, the operator, who is in the tower, opens the elevator, holds the length and then stows the upper end of the same length in racks (Figure I.4c).

This process continues until we reach the drill collars, which are also stored vertically in threes. The tripping in in which the drill string is run in the hole is known as Run in the Hole (RIH), and it is identical to the POOH. It should be noted that during this operation, we can neither rotate the drill string nor circulate the mud (Nguyen, 1993).

5.4. Backreaming & reaming

Backreaming and reaming are crucial processes in drilling operations that help ensure efficient and effective hole creation. Backreaming involves enlarging the diameter of the hole by moving the drill bit in a reverse direction, removing any debris or obstructions and creating a clear path for the drilling operation. Reaming, on the other hand, refers to the process of enlarging an existing hole to a specific diameter using a reamer tool. This is often done to achieve a more precise borehole that meets the desired specifications. Reaming can enhance the hole's quality, improve the well's integrity, and ensure that casing or tubing can be properly installed. Both backreaming and reaming play vital roles in the drilling process, optimizing the efficiency and success of the overall operation.

Chapter 3 : Machine Learning and its implementation on python

1. Introduction

As the volume of data increases, human cognition becomes increasingly limited in its ability to decipher important information. This situation is particularly evident in the oil and gas drilling field, where numerous rig sensors generate vast amounts of real-time data. For example, drilling operations may produce data on many parameters such as the rate of penetration (ROP), revolutions per minute (RPM), weight on bit (WOB), flow rate, torque, and other parameters. Machine learning is a powerful tool that can be employed to extract real-time insights from such data, providing valuable assistance to drilling crews in decision-making and enhancing their performance. In this chapter, we will define machine learning and provide an overview of the theory behind it, as well as discuss some well-known machine learning algorithms..

2. Machine learning algorithms

2.1. What is machine learning

Machine learning is an application of artificial intelligence that allows computer systems to learn and improve from experience without being explicitly programmed(Leimbach, 1994). The process of learning starts with collecting data or observations, searching for patterns in the data, and making decisions based on these patterns (Edgar & Manz, 2017). The aim of machine learning is to enable computers to learn automatically without human intervention. Machine learning algorithms can be classified into four types: supervised, unsupervised, semi-supervised, and reinforcement learning (Omar et al., 2013).

2.2. Machine learning types

2.2.1. Supervised Learning

This is used in the situations of the dataset which consists of a target or outcome variable which can be predicted from a given set of predicators (Osisanwo et al., 2017). Further, the goal of this is to predict the output for the new data once the algorithm identifies the known data. In addition to that, a supervised learning algorithm has two further processes such as classification and regression. The most widely supervised learning algorithms are linear regression, logistic regression, random forest, gradient boosted trees, support vector machines (SVM), neural networks, decision trees, Naïve Bayes and nearest neighbor (Ghorbani et al., 2015). In classification algorithms, incoming or new data is labeled based on the past samples of data and algorithms are trained to recognize the certain types of objects then classify accordingly. This is mainly differentiating the data based on similar features. Furthermore, regression is used to identify the patterns and calculate the predictions based on continuous outcome.

2.2.2. Unsupervised Learning (descriptive models)

In unsupervised learning, it consists of only input data and corresponding results are unknown. The goal of this model is to detect the underline structure or distribution of the data set through learning more about the data (Dy & Brodley, 2004). Further, this has been grouped into two segments due to the complexity of the logic including clustering and association (Khanum, 2015)[8]. Discovering the inherent groups in the data set is performed in the clustering process. However, association rule is used to describe the large proportions of data in the association process. The most popular examples of unsupervised learning algorithms are K-Means clustering, t-SNE (t-Distributed Stochastic Neighbor embedding and PCA (Principal Component Analysis and Association Rule (Dy & Brodley, 2004).

2.2.3. Semi-supervised Learning

These algorithms represent a middle ground between supervised and unsupervised features. However, some models combine the features of both aspects (Mao et al., 2016). This is basically assigning the situations where the dataset consists of both labeled and unlabeled data for training.

2.2.4. Reinforcement Learning (RL)

This interacts with its environment by producing actions and discovers errors or rewards. The decisions are taken sequentially i. e. outcome depends on the state of the current input and the next input depends on the output of the previous input (Mao et al., 2016).

3. Naive Bayes classifier

3.1. What is a Naive Bayes classifier

Naïve Bayes is the simplest probabilistic classifier in machine learning, commonly used in predictive modeling. The predicator variables in the Naïve Bayes are conditionally independent of other features. Further, Naïve Bayes classifier works by correlating with tokens and calculating the probability using Bayes Theorem to predict the event occurrences. Although, this algorithm is simple in nature, this is often used in more sophisticated activities. This study will explore the drilling operation classification using Naïve Bayes algorithm.

Naïve Bayes consists of four forms of its implementations such as Gaussian Multinomial, Complement and Bernoulli Naïve Bayes (Rennie et al., 2003) . The Multinomial Naïve Bayes has been used in this study due to the classification of discrete features such as bit positions. However, Gaussian is used for continuous data as inputs whereas complement classifier estimates parameters of a category. But Bernoulli assumes the distribution of probability as Bernullian.

3.2. How Naive Bayes classifier works ?

In the context of a classification problem where $X = X_1, X_2, \cdots, X_n$ represents an observation containing n independent attributes and C_j is one of K classes, the goal of the Naive Bayes Classifier is to compute the conditional probabilities $P(C_j/X_1, X_2, \cdots, X_n)$ for each class using Bayes' theorem.

Bayes' theorem allows us to calculate the posterior probability of class C_j given attribute X of an object or an individual based on prior knowledge as follows :

$$P(C_j/X) = \frac{P(C_j)P(X/C_j)}{P(X)}$$

where $P(C_j)$ is the prior probability of C_j, $P(X)$ is the prior probability of X, or the marginal probability, $P(X/C_j)$ is the likelihood of C_j, which quantify how likely it is to observe the realization of X given that the class is C_j.

The assumption of independence of variables can be expressed as follows :

$$P\left(X/C_j\right) = \prod_{i=1}^{n} P(X_i/C_j)$$

Using this expression to replace the term for the posterior probability of class C_j, we obtain ;

$$P(C_j/X) = \frac{P\left(C_j\right) \prod_{i=1}^{n} P\left(X_i/C_j\right)}{P(X)}$$

After computing $P(C_j/X)$ for the K classes, observation X will be assigned to class C if and only if :

$$P(C/X) \geq P(C_j/X) \qquad for\ each \qquad 0 \leq j \leq 1$$

The evaluation of created naïve model is done using test data. Accuracy score indicates how often the classifier makes the correct predictions. Further, the ratio of correct predictions to the total predictions. Precision Score indicates the ratio of true positive predictions to the total number of positive predictions made by the classifier.

4. Support Vector Machine

4.1. What is SVM?

Support Vector Machines are a type of supervised machine learning algorithms that provides analysis of data for classification and regression analysis. While they can be used for regression, SVM is mostly used for classification. We carry out plotting in the n-dimensional space. The value of each feature is also the value of the specific coordinate. Then, we find the ideal hyperplane that differentiates between the two classes. These support vectors are the coordinate representations of individual observation. It is a frontier method for segregating the two classes (Data-flair, n.d.).

4.2. How does SVM work?

The basic principle behind the working of Support Vector Machines is to create a hyperplane that separates the dataset into classes. To illustrate this concept, let us consider a scenario in which the task is to classify red triangles from blue circles. In this scenario, the objective is to create a line that classifies the data into two classes, creating a distinction between red triangles and blue circles (Data-flair, n.d.).

While one can hypothesize a clear line that separates the two classes, there can be many lines that can do this job. Therefore, there is not a single line that you can agree on which you can perform this task. Let us visualize some of the lines that can differentiate between the two classes as follows (Data-flair, n.d.).

In the above visualizations, we have a green line and a red line. Which one do you think would better differentiate the data into two classes? If you choose the red line, then it is the ideal line that partitions the two classes properly. However, we still have not concretized the fact that it is the universal line that would classify our data most efficiently (Data-flair, n.d.).

The green line cannot be the ideal line as it lies too close to the red class. Therefore, it does not provide a proper generalization which is our end goal. According to SVM, we have to find the points that lie closest to both the classes. These points are known as support vectors. In the next step, we find the proximity between our dividing plane and the support vectors. The distance between the points and the dividing line is known as the margin. The aim of an SVM algorithm is to maximize this very margin. When the margin reaches its maximum, the hyperplane becomes the optimal one (Data-flair, n.d.).

The SVM model tries to enlarge the distance between the two classes by creating a well-defined decision boundary. In the above case, our hyperplane divided the data. While our data was in 2 dimensions, the hyperplane was of 1 dimension. For higher dimensions, say, an n-dimensional Euclidean Space, we have an n-1 dimensional subset that divides the space into two disconnected components (Data-flair, n.d.).

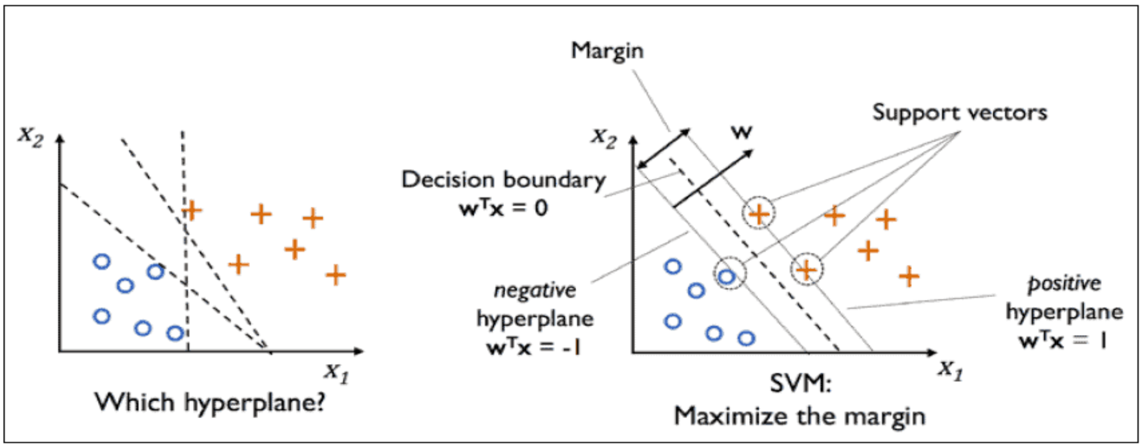

Figure II.1. SVM as Maximum Margin Classifier

4.3. Implementation of SVM in Python

After understanding the basics of SVM, we will explain in this section the implementation in Python. The latter is very simple with the SVM package of Scikit Learn, as described below:

a. Data importation

```
import pandas as pd
```

```
data = pd.read_csv(r'C:\file location\filename.csv')
```

b. Splitting the training and test datasets

```
from sklearn.model_selection import train_test_split

training_set, test_set = train_test_split(data, test_size = 0.2, random_state = 1)

X_train = training_set.iloc[:,0:2].values

Y_train = training_set.iloc[:,2].values

X_test = test_set.iloc[:,0:2].values

Y_test = test_set.iloc[:,2].values
```

c. SVM model creation

```
from sklearn.svm import SVC

classifier = SVC(kernel='rbf', random_state = 1)

classifier.fit(X_train,Y_train)X_test = test_set.iloc[:,0:2].values
```

d. Model evaluation

```
Y_pred = classifier.predict(X_test)

accuracy = accuracy_score(y_test, y_pred)
```

5. Decision Tree

5.1. What is the Decision Tree?

A decision tree is a machine learning method used to explain a target value (y) based on a set of input variables (X) and visualized in a tree-like structure. The target values of y can be either discrete (classification tree) or continuous (regression tree).

The main components of a decision tree are nodes and branches.

- The root node, also known as the decision node, represents the first node that will partition all the observations of the data set into two or more mutually exclusive subsets.
- The internal nodes, also known as decision nodes, are the nodes located between the root node and the leaf node, labeled by decision rules corresponding to the attributes.
- The leaf node, also known as the output node, represents the final result of a combination of decision rules or events.

- The branches represent the possible responses to decision rules or occurrences that emanate from internal nodes and the root node. To predict the label of an observation, the decision tree model is formed by using a hierarchy of branches.

Each path from the root of the tree to the leaf represents a successive series of decision rules. These decision tree paths can also be represented in the form of "if-then" rules. For example, "if condition 1 and condition 2 and... condition j occur, then the result k occurs."

5.2. How decision trees work ?

Consider a training data set S characterized by n class C_1, C_2, ..., C_n and p input variables or attributes X_1, X_2, ..., X_p. There are many decision trees that can be constructed from the training set S. To find the most accurate tree among these trees, decision tree construction algorithms typically use a strategy that first tries to identify which input variable can provide the best partitioning of the data observations from the root node.

The algorithm then recursively partitions the decision data to define the decision tree structure, selecting at each decision node the separating variables, from the most to the least discriminative. This separation process continues until the stopping criteria are satisfied.

The concept used for the selection of the decision attribute X that better discriminates the data set depends on the decision tree construction algorithm used. The C4.5 and ID3 algorithms, for example, respectively employ information gain and gain ratio, two division criteria based on the same concept called entropy. Meanwhile, the CART algorithm is based on another concept called the Gini index.

Chapter 4 : Development of an intelligent system for ILT measurement

1. Introduction

The drilling of wells in the field is a complex and challenging operation that involves numerous technical and logistical factors. However, one significant problem that frequently arises during drilling operations is the issue of lost time due to Invisible Lost Time (ILT). ILT can be highly frustrating and impede the progress of drilling activities, causing delays and increased costs for oil and gas companies.

This chapter focuses on the development of an intelligent system to measure and mitigate invisible time during drilling operations. To achieve this objective, we employ advanced techniques and methodologies. Firstly, we create a rig state for automatic drilling operations classification using support vector machine (SVM) algorithms. This automated classification enables a real time identification of drilling rig states.

Next, the built SVM model was applied to assess key performance indicators (KPIs) for drilling connection operations in two different wells well-A and well-B. Lastly, we focus on determining Invisible Lost Time (ILT). By combining these methods, our intelligent system aims to accurately measure, analyze, and mitigate invisible time in oil and gas drilling operations.

2. Drilling operation classification using SVM model

2.1. Training SVM model

In this section, we discuss the process of training an SVM model for rig state creation. The SVM model is an effective machine learning algorithm used to classify drilling operations accurately. To train the model, we generated a data sample, which provides valuable insights into how drilling operations should be classified.

Table 4.1 represent the data that was generated for training the SVM models. As we can see in the table, the data sample consists of six columns, with the first five columns representing the inputs and the last column representing the classes or outputs, which correspond to different drilling operations.

The drilling operations included in the classes are drilling, in slips, reaming, pull out of the hole (POOH), run in the hole (RIH), backreaming, and circulation. The inputs in the data sample include RPM (rotations per minute), torque, flow in values, and the state of the hook position, hook load, and bit position. These parameters play a crucial role in determining the current state of the drilling operation and aid in classifying the rig state accurately.

By utilizing this training data, we can create an accurate classification model that takes into account the various inputs and their corresponding drilling operation classes. The SVM model learns from this data sample, allowing it to generalize patterns and make predictions based on new input values.

It is important to highlight the value of this training data, as it serves as the foundation for creating a reliable and robust classification model. By training the SVM model with this data, we can develop a system that accurately identifies and classifies different drilling operations, providing insights into the rig state during drilling operations.

Bitpos	hookload	blockpos	rpm	Torque	flowin	Rigstate
Nul	val2	descent	>10	>100	>10	Drilling
positive	val2	descent	nul	nul	nul	RIH
positive	val2	Static	nul	nul	nul	Stationary
positive	val2	ascend	nul	nul	nul	POOH
positive	val2	ascend	>10	nul	>10	Backreaming
positive	val2	stop	>10	nul	>10	Circulation
positive	val2	descent	>10	nul	>10	Reaming
negative	val2	descent	nul	nul	nul	Error
positive	val1	descent	nul	>100	nul	Inslips

Tableau 2.1. *Sample of data used for rig state creation*

2.2. Testing the SVM model

In this subsection, we will discuss the process of testing the SVM model, which involves evaluating its performance on new data samples. By assessing the model's accuracy and effectiveness, we can ensure its reliability in classifying drilling operations in real-time.

The created SVM model was implemented to detect drilling operations based on drilling data. The developed SVM model was applied in two wells well-A et well-B to test its ability to correctly classify the rig state in real-time while drilling three sections, including the 26, 16’’ and the 8½.

The testing process was implemented using Python, and the steps followed for testing the trained SVM model are ordered as follow :

1. Import or install the necessary libraries.
2. Load the trained SVM model.
3. Prepare your test data (Make sure your test data is preprocessed and formatted similarly to the training data).
4. Make predictions on the test data.
5. Evaluate the accuracy of the model.

During this phase we noticed that the training data consisted of categorical information, while the testing data primarily contains numerical values related to drilling operations. To address this issue, we developed a coding block using Python to process the drilling data before passing it to the trained rig state model for identification.

Figure 4.1 illustrates the flowchart of this coding block. It begins by receiving input parameters such as the values of drill bit value, depth, previous block position, actual block position, …. etc. The coding block then processes these values to generate the drill bit state, determining whether it is on the bottom or not. For instance, if the drill bit position is lower than the depth, it indicates that the drill bit is not on the bottom; otherwise, it is considered on the bottom. Similarly, the block position state is derived by comparing the previous and current positions of the block. This comparison allows us to determine whether the block position has changed or remained the same. By employing this coding block, we preprocess the testing data to transform numerical inputs into categorical representations that fits with the training data. This ensures compatibility with the trained rig state model, enabling accurate identification of the rig state during drilling operations.

The results obtained after testing the rig state model demonstrate a high level of accuracy in identifying rig states from the drilling data. This validates the reliability and effectiveness of the model in real-time detection of Invisible Lost Time (ILT) in the next sections. The ability to rely on the trained rig state model to detect ILT in real time has significant implications for the oil and gas drilling industry. It empowers operators and decision-makers to promptly address inefficiencies, optimize drilling processes, and minimize downtime, ultimately leading to improved operational efficiency .

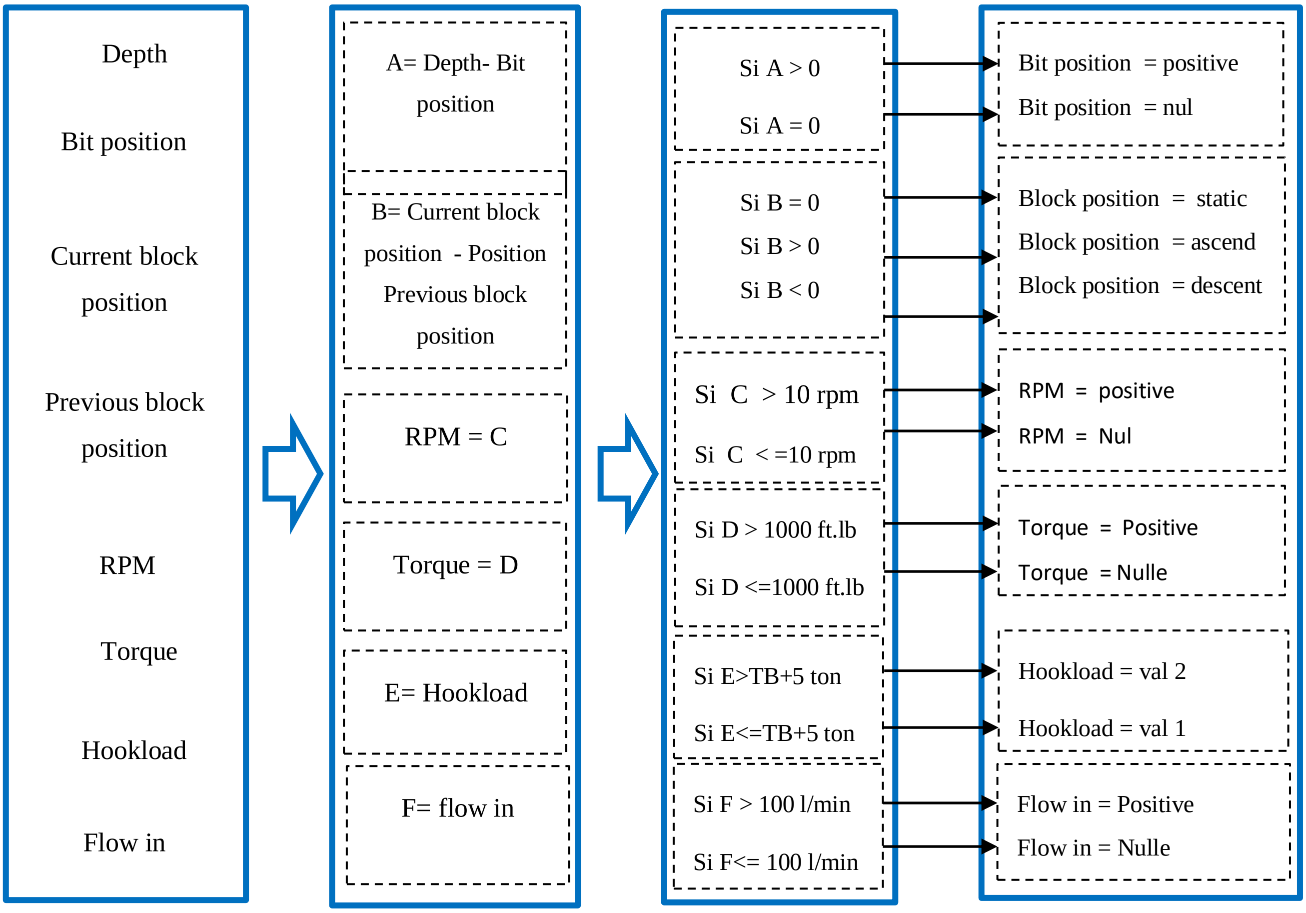

Figure.IV.1. *Flowchart used for processing the drilling data*

3. Assessment of the Key performance indicator for connection operation

3.1. Drilling connection time assessment

In this section, we describe how we applied the created SVM drilling operation classification model to derive Key Performance Indicators (KPIs) for drilling connection time. We gathered data from two wells, Well A and Well B, and utilized the trained rig state model to label this data and identify the connection operations, which corresponded to the rig state "in slips."

Using Python, we developed a code that computed the total time taken to establish each connection based on the data labelled by the rig state. This enabled us to determine the drilling connections performed by both the daily and night crews in Well A and Well B during the drilling of three phases 26, 16’’ and the 8½, as well as the time consumed for each connection. This assessment allowed us to monitor and assess the performance of each crew in the two wells.

Figure.IV.2 vrepresents histograms illustrating the total time taken for drilling connections by the daily and night crews in both wells. The histograms reveal that Crews A and B in Well A, as well as Crew B in Well B, were able to establish most of the connections within a duration of 3 to 6 minutes. Notably, Crew A in Well B demonstrated exceptional efficiency and performance, completing the majority of connections within just 2 minutes.

3.2. Outliers detection

In this section, we proceeded to compute the KPI for drilling connection. To achieve this, we conducted data analysis of the connection time in an Excel spreadsheet. To ensure the accuracy and reliability of the computed Key Performance Indicator (KPI) for drilling connection time, we took measures to eliminate outliers from the data.

Outliers are data points that deviate significantly from the overall pattern of the dataset and can distort the results of statistical analyses. These outliers could represent events such as extending the drilling cable or conducting well checks, or other operations which may occur during the "in slips" rig state.

To identify outliers, we employed the Interquartile Range (IQR) method. The IQR is a measure of statistical dispersion, representing the range between the first quartile (Q1) and the third quartile (Q3). Q1 is the value below which 25% of the data falls, while Q3 is the value below which 75% of the data falls. The following equation were used to calculate the upper limit and lower limit for identifying outliers.

$$IQR = Q3 - Q1$$
$$Upper\ limit\ = Q3 + 1.5\ IQR$$
$$Lower\ limit\ = Q1 - 1.5\ IQR$$

Any data points that fell below the lower limit or above the upper limit were considered outliers and subsequently discarded. This step ensured that the outliers would not influence the findings of the statistical study performed to determine the drilling connection KPI.

By eliminating the outliers, we obtained a cleaner and more representative dataset, enabling a more accurate computation of the KPI. This process enhanced the reliability of the statistical analysis and ensured that the KPI reflected the typical and meaningful drilling connection times.

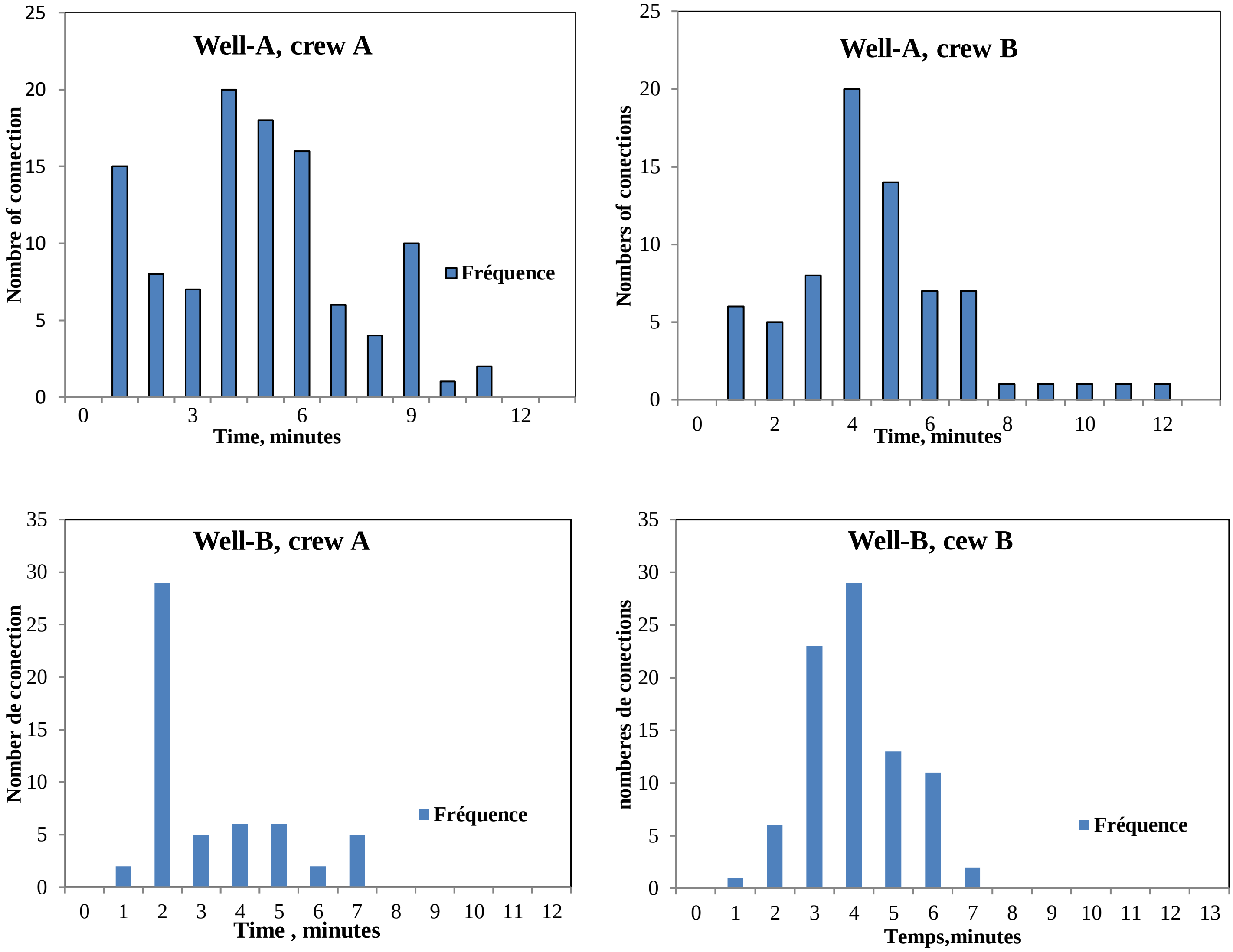

***Figure*.IV.2.** *Histogram distribution of the drilling connection for the crews A and B in the wells A and B*

3.3. KPI assessment

Once the outliers were eliminated, we calculated the KPI for drilling connection time. Various approaches exist for computing the KPI, including the average, the P50 method, and the "best of the best" approach. In this study, we considered the KPI for drilling connection as the average time of all drilling connection times. After applying the average approach, the computed KPI was found to be 4 minutes and 13 seconds.

This KPI serves as a valuable metric for evaluating the efficiency and performance of drilling connection operations. It provides insights into the average time taken to establish connections, enabling operators to identify areas for improvement and optimize the overall drilling process.

4. Determination of the ILT (invisible lost time)

In this section, we focused on measuring the Invisible Lost Time (ILT) associated with drilling operations. ILT refers to the time lost due to inefficiencies within the drilling crew, we cannot asses this lost time and it is considered as invisible because is it not depicted in the drilling reports. To quantify ILT, we conducted a statistical study using an Excel spreadsheet.

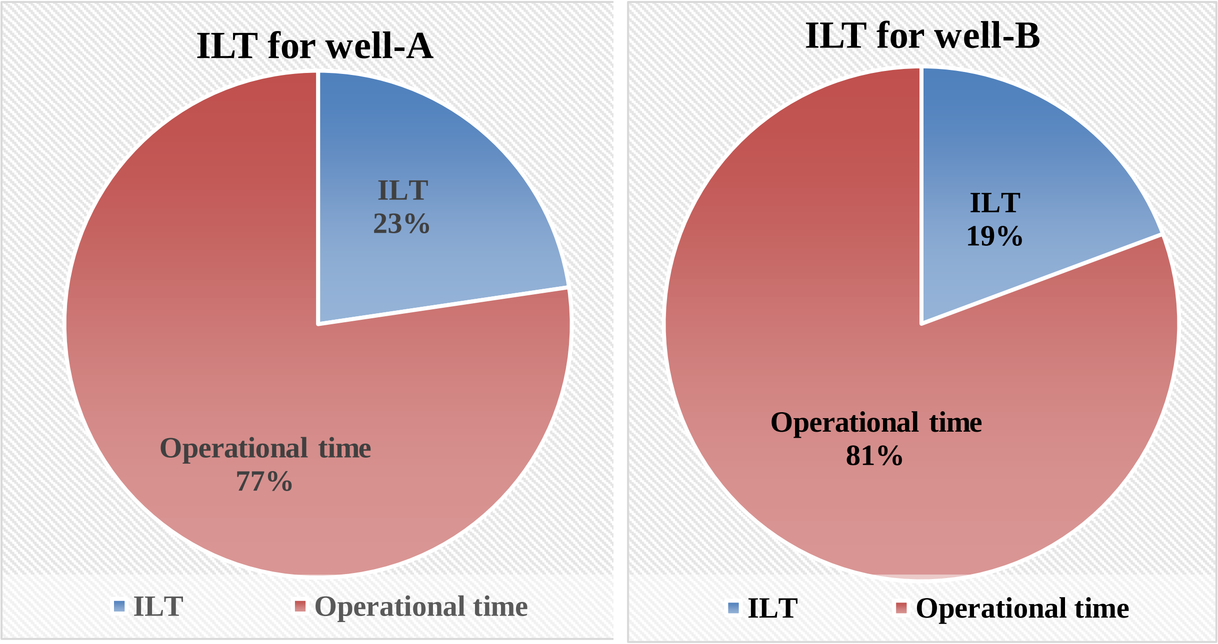

Figure.IV.3. *Pie chart of ILT in the well A and B*

This analysis involved examining the total connections established by Crew A and Crew B in both wells, along with the corresponding time taken to complete each connection. Any time consumed for establishing drilling connections that is greater than the connection KPI was considered as invisible lost time, and thus the total ILT were measured. Figure 4.3 and Table 4.2 provide visual representations of ILT in the two wells. These illustrate the total ILT and its proportion to the total time expended on drilling connections.

Based on Figure 3 and Table 2, it can be observed that for well-A the total ILT is 2 hours and 53 minutes, which is equivalent to 22.67 % of the total time dedicated to establishing drilling connections. For well-B, the total ILT is 1 hours and 44 minutes, which is equivalent to 19.35 % of the total time dedicated to establishing drilling connections.

	Well A	Well B
Number of connections	163	145
Total time of connection	12 hours and 43 minutes	8 hours and 59 minutes
Invisible lost time	2 hours and 53 minutes	1 hours and 44 minutes
Fraction of ILT	22.67 %	19.35 %

Tableau 4.1. *ILT for well-A and well-B*

5. Conclusion

Using Python for analysing lost time invisible in drilling operations provides a data-driven approach to understand the causes, quantify the impact, compare performance, identify trends, and make informed decisions for optimizing drilling processes. This analysis can lead to improved efficiency, reduced costs, and enhanced productivity in drilling operations. Calculating the lost time invisible in drilling operations is crucial for improving efficiency and productivity in the drilling industry. By identifying and addressing the causes of lost time, operators can minimize delays, optimize operations, and reduce costs.

In conclusion, calculating the lost time invisible in drilling operations provides valuable insights into the causes and extent of inefficiencies. This analysis helps in identifying areas for improvement, setting performance targets, and implementing strategies to enhance operational efficiency and productivity. By addressing lost time, drilling companies can optimize their operations, reduce costs, and improve overall performance.

General Conclusion

In conclusion, there is a lot of opportunity to increase productivity and efficiency in the drilling business by creating an intelligent system for calculating hidden wasted time during drilling operations. Such a solution can assist operators and decision-makers in optimizing drilling procedures and lowering expensive downtime by precisely detecting and measuring the hidden causes of lost time and the development of an intelligent system for measuring invisible lost time during drilling operations using Python SVM (Support Vector Machine) can offer valuable insights and predictive capabilities to optimize drilling processes.

Advanced technologies including artificial intelligence (AI), machine learning, and data analytics would probably be included into the intelligent system. It will gather and evaluate critical metrics associated with drilling operations using real-time data from numerous sensors and monitoring devices mounted in drilling rigs.

By applying AI and machine learning algorithms, the system can identify patterns and correlations in the data, enabling it to detect and classify different types of invisible lost time, such as equipment failures, inefficient procedures, or unexpected events. It would also consider factors like weather conditions, geology, and operational constraints to provide a comprehensive assessment of lost time.

In summary, there is considerable potential to increase drilling efficiency and decrease downtime by creating an intelligent system for detecting hidden lost time during drilling operations using Python SVM. The technology may assist operators in identifying and addressing inefficiencies, optimizing drilling procedures, and eventually improving overall performance by utilizing the capabilities of machine learning, real-time data analysis, and predictive modelling.

References

Bello, O., Teodoriu, C., Yaqoob, T., Oppelt, J., Holzmann, J., & Obiwanne, A. (2016). Application of artificial intelligence techniques in drilling system design and operations: a state of the art review and future research pathways. *SPE Nigeria Annual International Conference and Exhibition*.

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

CHERIFI, M. (2012). *Etude et maintenance de treuil de forage OIL WELL 840E*. Memoire de fin d'étude, Universte Hassiba Benboualide, chlef.

Coley, C. (2019). Building a Rig State Classifier Using Supervised Machine Learning to Support Invisible Lost Time Analysis. *SPE/IADC International Drilling Conference and Exhibition*.

Data-flair. (n.d.). *Support Vector Machines Tutorial – Learn to implement SVM in Python*.

Dy, J. G., & Brodley, C. E. (2004). Feature selection for unsupervised learning. *Journal of Machine Learning Research*, *5*(Aug), 845–889.

Edgar, T. W., & Manz, D. O. (2017). Chapter 6-machine learning. *Research Methods for Cyber Security*, 153–173.

Ghorbani, A., Steinhilber, G., Markus, D., & Maas, U. (2015). A PDF projection method: A pressure algorithm for stand-alone transported PDFs. *Combustion Theory and Modelling*, *19*(2), 188–222.

Leimbach, M. (1994). Expertsystem-model coupling within the framework of an ecological advisory system. *Ecological Modelling*, *75*, 589–600.

Mao, H., Alizadeh, M., Menache, I., & Kandula, S. (2016). Resource management with deep reinforcement learning. *Proceedings of the 15th ACM Workshop on Hot Topics in Networks*, 50–56.

Nguyen, J.-P. (1993). *Forage (Le)*. Editions OPHRYS.

Omar, S., Ngadi, A., & Jebur, H. H. (2013). Machine learning techniques for anomaly detection: an overview. *International Journal of Computer Applications*, *79*(2).

References

Osisanwo, F. Y., Akinsola, J. E. T., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., & Akinjobi, J. (2017). Supervised machine learning algorithms: classification and comparison. *International Journal of Computer Trends and Technology (IJCTT)*, *48*(3), 128–138.

Rennie, J. D., Shih, L., Teevan, J., & Karger, D. R. (2003). Tackling the poor assumptions of naive bayes text classifiers. *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, 616–623.

Sadanandan, N. M. (2014). *Enhancing Directional Drilling Using Wired Drill Pipe Telemetry*. Teesside University.

Vempati, C., Moss, W., Townsend, T., Schandorf, C., & Evans, K. (2020). Lessons Learned From In-Bit Weight and Torque Measurement. *SPE Annual Technical Conference and Exhibition*.

Yin, Q., Yang, J., Hou, X., Tyagi, M., Zhou, X., Cao, B., Sun, T., Li, L., & Xu, D. (2020). Drilling performance improvement in offshore batch wells based on rig state classification using machine learning. *Journal of Petroleum Science and Engineering*, *192*, 107306.