PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
Ministry Of Higher Education and Scientific Research

Amar Thelidji University- Laghouat

FACULTY: TECHNOLOGY

DEPARTMENT: Process Engineering

Master's dissertation

Defended by: Farah Fatmazohra AZZAOUI

DOMAIN: Science and Technology

FIELD: Process Engineering

SPECIALTY: Hydrocarbons

OPTION: Gas Natural Engineering

Theme

Exploration of real-time drilling data for optimization using machine learning

Supervisory and Eaxamining Committee :

Name and First Name	**Grade**	**Quality**
ABDELMOUIZ Ahmed	MCB	Chair of defence
YOUCEFI Riad	MAB	Examiner
HADJADJ Asma	MCB	Supervisor
HADJKOUIDER Mohammed	MCA	Co-Supervisor

Promotion: JUIN 2022

Acknowledgements

First and foremost, praises and thanks to ALLAH, the Almighty, for His showers of blessings throughout my research work to complete the research successfully,

I wish to express my sincere gratitude to my supervisor Dr. Asma HADJADJ, for her enthusiasm, patience, insightful comments, helpful information, it was a great privilege and honor to work and study under her guidance. I would also like to thank her for friendship, empathy, and great sense of humor practical advice, I could not have imagined having a better supervisor in my study.

Special thanks for Dr Mohammed HADJKOUIDER Professor at AHMED DRAYA University in ADRAR for his unceasing ideas that have helped me tremendously at all times in my research and writing of this thesis. His immense knowledge, profound experience and professional expertise in Data Quality Control has enabled me to complete this research successfully. I am extremely grateful for what he has offered me.

I also wish to express my sincere thanks to the University of AMAR TELIDJI for accepting me into the Master's degree, without that acceptation, this work couldn't never start.
Also, I thank all professors and staff of the process Engineering Department and everyone who helped me to achieve this dissertation.

Dedication

I dedicate this dissertation work to my family and my friends. A special feeling of gratitude to the secret of my happiness “my mother” and “my dear grandmother”, I hope that you are proud of me in addition to my beloved father. My sisters Zahira and Houaria have never left my side and are very special. I also offer this dissertation to many of my friends who have supported me all over the road. I will always appreciate all they have done.

Table of contents

List of Abbreviations

ML	Machine Learning
MWD	Measure while Drilling
NPT	Non-Productive Time
RF	Random Forest
MWD	Measure while Drilling
NPT	Non-Productive Time
RF	Random Forest
ROP	Rate Of Penetration
RPM	Rotation Per Minute
WOB	Weight on Bit
LWD	Logging While Drilling
ECD	Equivalent Circulating Density
TDS	Top Drive System
E&P	Petroleum Exploration and Production
AI	Artificial Intelligence
NLP	Natural Language Processing
KNN	k-Nearest Neighbors
SVM	Support Vector Machine

List of figures

List of tables

Introduction

Introduction

Drilling optimization is the “process of designing equipment and selecting operational parameters to minimize the well drilling cost”. Thus, the rate of penetration (ROP), defined as the volume of rock removed expressed as depth per time unit, is a key metric to measure drilling performance.

Although a high ROP is considered as a good metric to measure drilling efficiency and performance, drilling faster can affect cutting transport and lead to the bore hole instability issues and poor hole cleaning. It varies from the rock type being drilled, so it can give an idea of bit wear. Therefore, oil and gas companies look for high ROP values while operating in safety standards.

Generally, the ROP optimization endeavour depends on dynamic and static drilling parameters. Dynamic parameters can be the controllable weight on bit (WOB), rotary speed (RPM), torque [T) or the uncontrollable ones, whether the driller can alter it manually during operations or not. Static parameters include formation properties such as compressive strength and formation pressure among other things.

The motivation of this work is to implement a tool (numerical code) that can automatically improve the drilling efficiency by maximizing the ROP. This code will provide the dynamic controllable parameters to achieve the desirable ROP and therefore save drilling costs.

The goal is therefore to maximize the Rate of Penetration and in order to achieve a good modelling, it is important to have an accurate model to predict the values to be optimized

The project can be summarized in the following tasks:

- Clean and process all dataset.
- Realize Machine Learning Models.
- Implement Machine Learning Algorithms in the dataset.
- Evaluate Machine Learning Models Performance.

Chapter I:

Generality of drilling

I.1 Drilling Optimization History

The concept of rotary drilling originated in the beginning of the year 1900. The development of rotary drilling can be divided into four distinct periods: conception period 1900 to 1920, development period 1920 to 1950, scientific period 1950 to 1970, and automation period which began in 1970. The conception period the rotary drilling principle marked the usage of cementing methods, rotary bits, drilling fluids and casing installations. In 1950s the scientific period took place with expansion in drilling research, better understanding of the hydraulic principles, significant improvements in bit technology, improved drilling fluid technology and most important of all optimized drilling. After 1970s rigs with full automation systems, closed-loop computer systems, with ability to control the drilling variables started to operate in oil and gas fields.[1]

Figure 1 gives the time line of drilling optimization history. One of the first attempts for the drilling optimization purpose was presented in the study of Graham and Muench in 1959.[2] They analytically evaluated the weight on bit and rotary speed combinations to derive empirical mathematical expressions for bit life expectancy and for drilling rate as a function of depth, rotary speed, and bit weight. In 1963 Galle and Woods [3] produced graphs and procedures for field applications to determine the best combination of drilling parameters. One of the most important drilling optimization studies performed was in 1974 by Bourgoyne and Young [4]. They proposed the use of a linear drilling penetration rate model and performed multiple regression analysis to select the optimized drilling parameters. They used minimum cost formula, showing that maximum rate of penetration may coincide with minimum cost approach if the technical limitations were ignored.

In the mid-1980s operator companies developed techniques of drilling optimization in which their field personnel could perform optimization at the site referring to the graph templates and equations. In 1990s different drilling planning approaches were brought to surface. New techniques identified the best possible well construction performances [5].

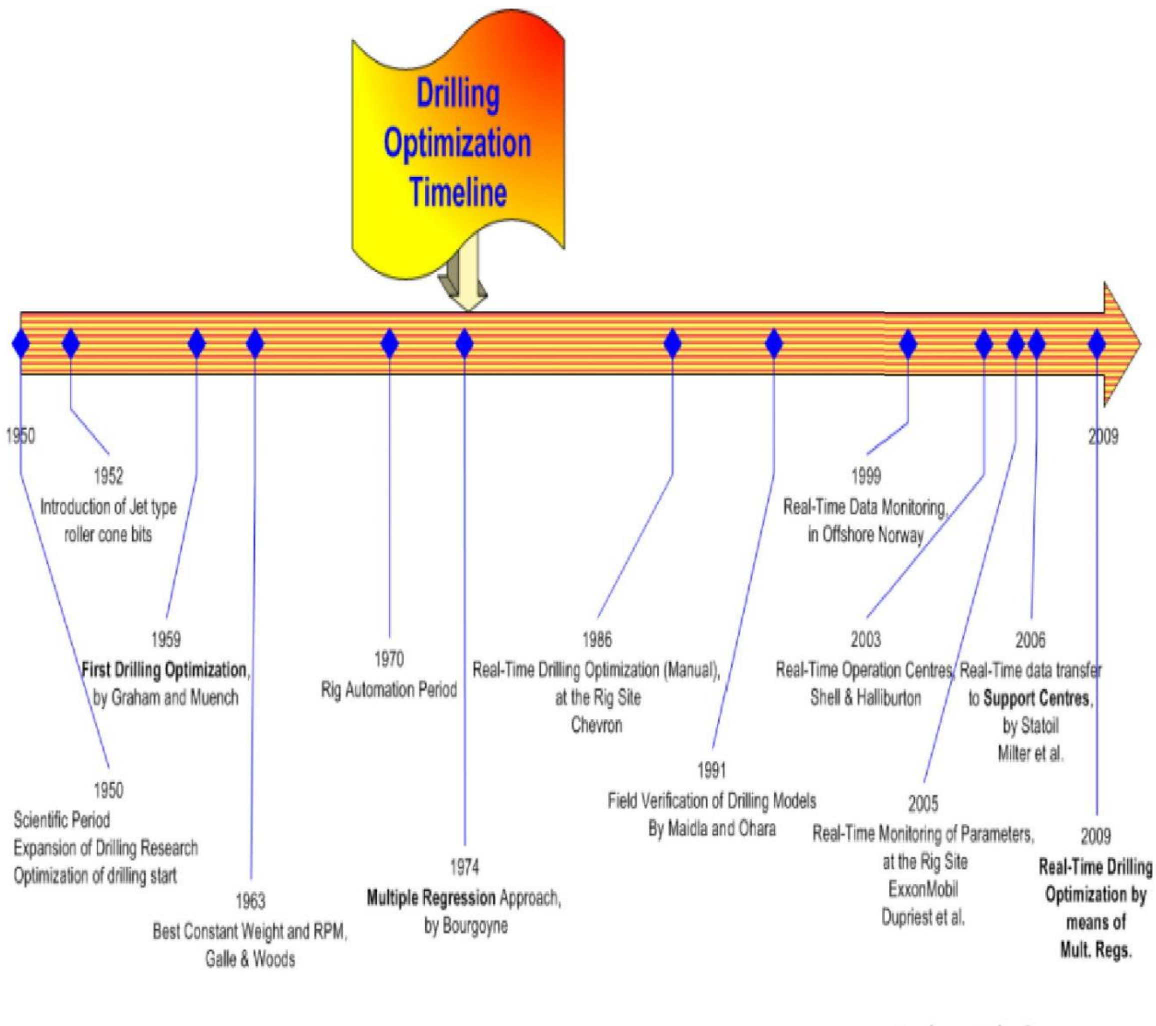


Figure 1: The time line of drilling optimization history [1]

I.2 rotary drilling process:

In rotary drilling the rock is bored using a cutting tool called the bit, which is rotated and simultaneously forced against the rock at the bottom of the hole by a drill string consisting of hollow steel pipes of circular section screwed together. The cuttings produced by the bit are transported up to the surface by a drilling fluid, usually a liquid (mud or water), or else a gas or foam, circulated in the pipes down to the bit and thence to the surface [6].

The rotation is transmitted to the bit from the surface by a device called the rotary table or, in the modern rigs, by a top drive motor with the rotary table as backup, additional rotation can be added by downhole motors located directly above the bit. After having drilled a certain length of hole, in order to guarantee its stability, it has to be cased with steel pipes, called casings, joined together by threaded sleeves. The space between the casing and the hole is then filled with cement slurry to ensure a hydraulic and mechanical seal [6].

The final depth of the well is accomplished by drilling holes of decreasing diameter, successively protected by casings, likewise of decreasing diameter, producing a structure made up of concentric tubular elements [6].

I.3 drilling rig:

Is designed to drill by applying rotation to the drilling head driven by hydraulic oil pressure, which eventually drives the drilling string to rotate. Bits like roller, alloyed bit and short bits are used [7].

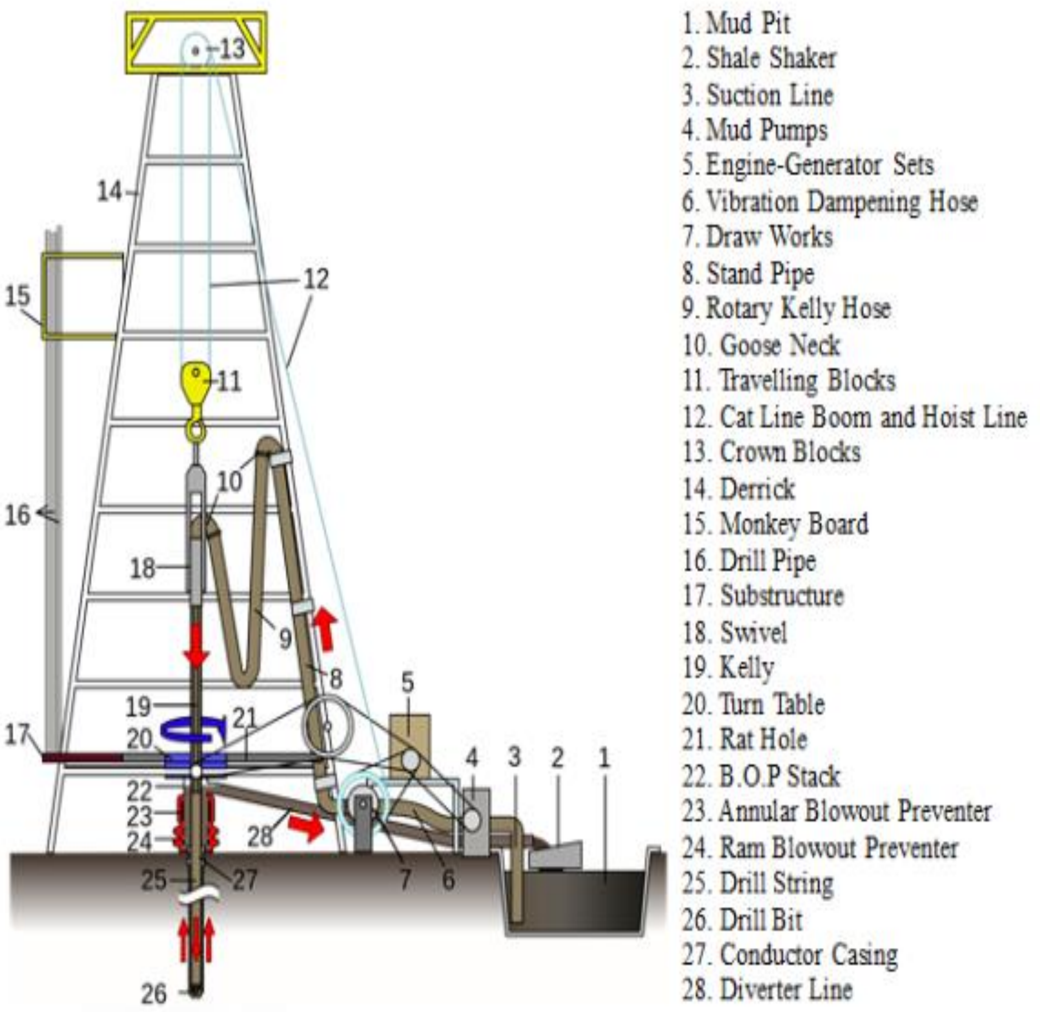


Figure 2: Rotary drilling rig [8]

- **Drilling string:** The drill string means the tubular and the accessories used to run the drill bit to the bottom. The drill string is composed of drill pipe, heavy weight drill pipe, drill collars and other components like stabilizers and drilling jars.

There are many functions of the drill string among them:

- Suspend the drill bit.

- Transmit the rotary motion from the Kelly or top drive to the drill bit.

- Provide a flow path to circulate drilling fluids.[9]

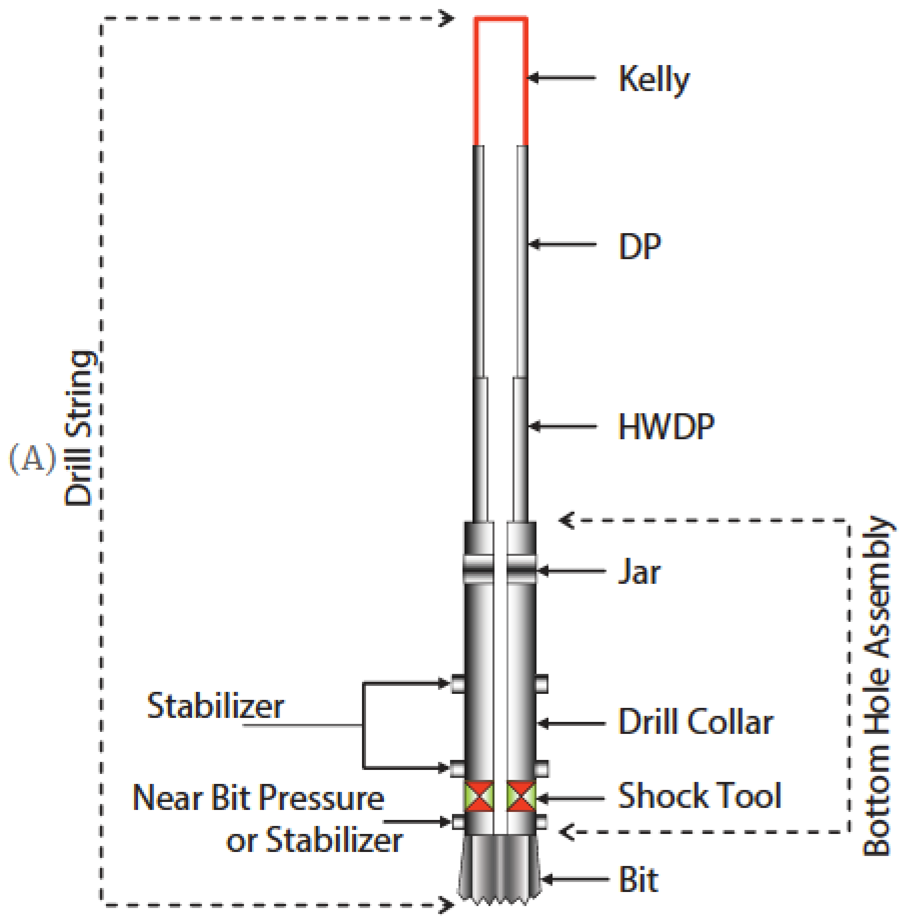


Figure 3: Components of the drill string [10]

- **Drill bit**: The cutting or boring element used in drilling oil and gas wells. Most bits used in rotary drilling are roller-cone bits. The bit consists of the cutting elements and the circulating element. The circulating element permits the passage of drilling fluid and uses the hydraulic force of the fluid stream to improve drilling rates.[11]

Figure 4: Drill bit [11]

- **Drill Collars:** A heavy, thick-walled tube, usually steel, used between the drill pipe and the bit in the drill stem. It is used to put weight on the bit so that the bit can drill.[11]

Figure 5: Drill collars DC [11]

- **Drill Pipe**: The heavy seamless tubing used to rotate the bit and circulate the drilling fluid. Joints of pipe 30 feet long are coupled together with tool joints [11].

Figure 6: Drill pipe [11].

Kelly: The Kelly has a square or hexagonal cross-section and provides the rotation of the drillstring. Because the Kelly is made of high quality, treated steel, it is a flashy part of the drillstring. Thus, to prevent the Kelly from excessive wear caused by making and breaking connections, a Kelly sub is mounted at the bottom end of it. To prevent backward flow of the mud in case of a kick, a Kelly cock providing a backflow restriction valve is often mounted between Kelly and swivel [12]

Figure 7: Kelly [13]

I.4 mud circulating system:

The drilling fluid circulating system is like a close loop electric circuit through which drilling fluid can travel from the surface to all the way downhole and back to its initial point (mud pit).[14]

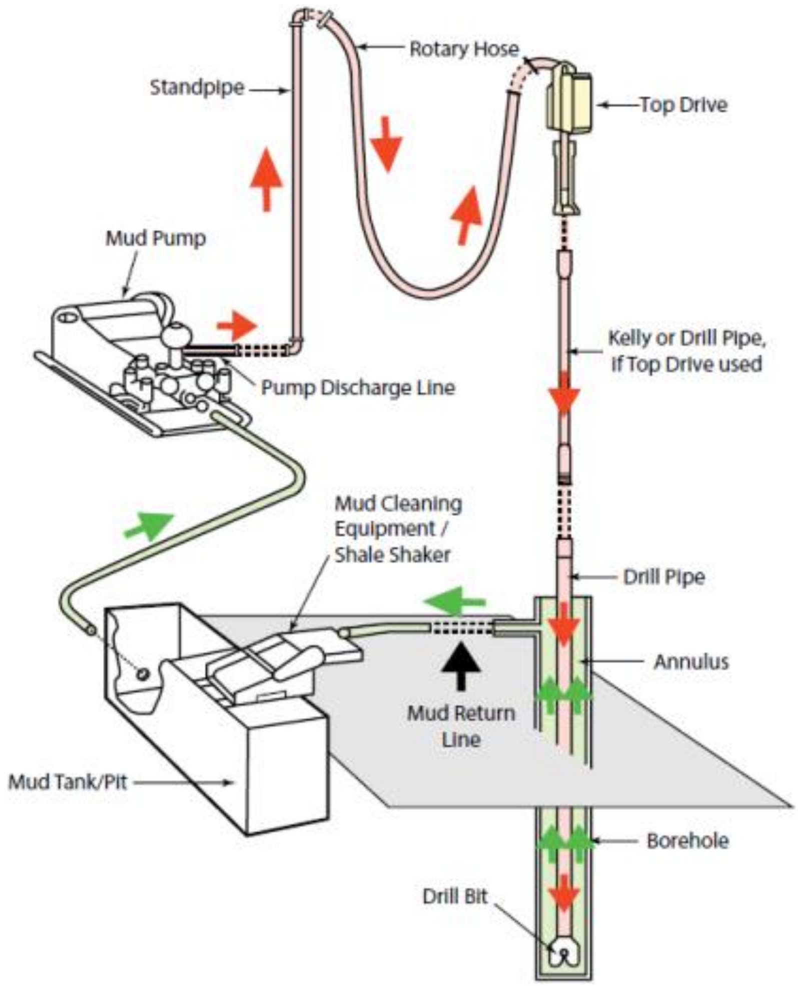


Figure 8:Mud circulating system [14]

Drilling fluid goes from the mud pits to main mud pump, and then major components including surface piping, standpipe, Kelly hose, swivel, Kelly, drill pipe, drill collar, bit nozzles, the various annular, flow line, mud control equipment, mud tanks, and again the mud pit/mud pump (Figure I.8).

The rock cuttings must be removed from the borehole to allow drilling to proceed. This is done by pumping drilling fluid down the drill-string, through the bit and up the annulus.[15]

The cuttings are then separated from the mud, which is then recycled. The circulating system (i.e. drilling fluid) also enables to clean the hole of cuttings made by the bit; to exert a hydrostatic pressure sufficient to prevent formation fluids entering the borehole, and to maintain the stability of the hole by depositing a thin mud-cake on the sides of the hole.[14]

I.5 recording drill data (mud logging):

Real-time drilling parameters that may be compiled include: rate of penetration (ROP), pump rate (quantity of fluid being pumped), weight on bit, rotary speed, rotary torque, RPM (Rotation Per Minute), mud weight and mud viscosity. These informations are usually obtained from mud logging unit by attaching monitoring devices to the drilling rig equipment with a few exceptions such as the mud weight and mud viscosity which are measured by the derrick hand or the mud engineer.

- **A mud logging unit***:* is installed on the rig when geologic information must be retrieved on a timely basis. If the formations drilled are not well known or if a specific geologic horizon is targeted, mineralogical data from cuttings brought to the surface with the mud might be essential [15]. mud logging units can discover and evaluate oil & gas reservoirs rapidly, provide complete drilling parameters and monitor toxic gases, thus ensuring drilling safety, enhancing drilling efficiency and reducing operation cost.[16]
- **Mud logging services:** is the creation of a detailed record (well log) of a borehole operation by examining the cuttings of rock brought to the surface by the circulating drilling medium.[17] Is provides subsurface geological information while drilling a well. Mud logging examines and analyzes geological information contained in formation cutting and drilling mud, to determine if oil and gas are encountered during well drilling. Mud logging also provides critical safety function such as determine pore pressure, kick control and ambient gas monitoring.[18]

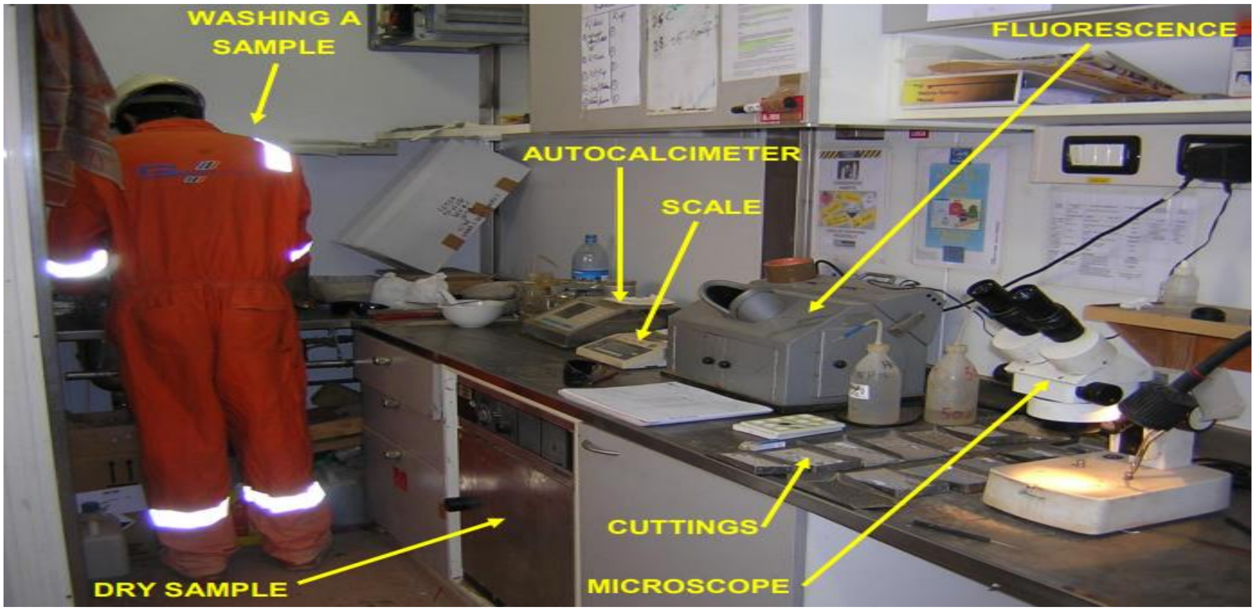


Figure 9: Mud logging unit [19]

Mud loggers connect various sensors to the drilling apparatus and install specialized equipment to monitor or "log" drill activity. This can be physically and mentally challenging, especially when having to be done during drilling activity. Much of the equipment will require precise calibration or alignment by the mud logger to provide accurate readings.

Mud logging requires a good deal of diligence and attention. Sampling the drilled cuttings must be performed at predetermined intervals, and can be difficult during rapid drilling [20].

Mud logging totally depends upon the mud circulation principle in well which is going to drill (figure 10).

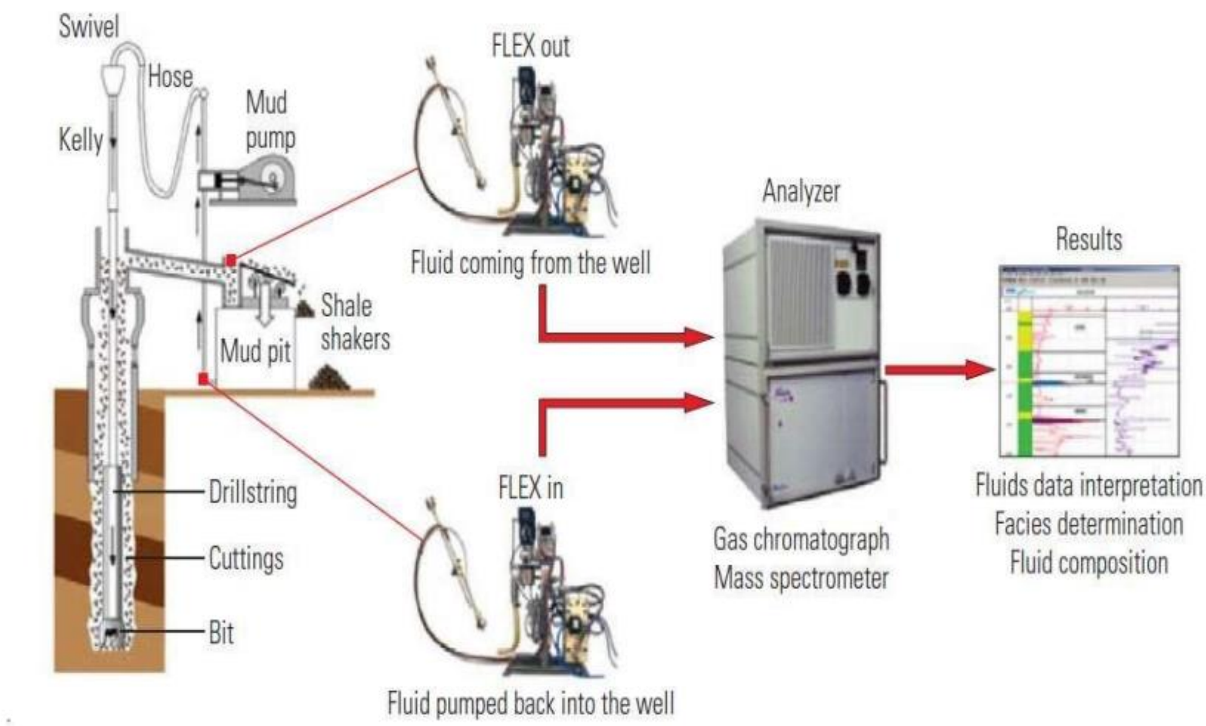


Figure 10: Record drill data process [21]

I.6 drilling parameters:

The parameters recorded for drilling optimization are critically important to be representative of data they are meant to reflect. The brief description of the drilling parameters is deemed necessary to be explained.[22]

Depth: The value of depth, in other means the bit position is input in the MLU. The operator is the responsible for that; usually it is linked to the position of the block, by means of the sensors located at the crown block.[22]

Weight on bit (**WOB)**: It is the abbreviation for "Weight on Bit". It represents amount of weight applied onto the bit, that is then transferred to the formation which in turn is the energy created together with string speed that advances drill string. It is measured through the drilling line, usually by means of having attached a strain-gauge which measures the magnitude of the tension in the line itself, and gives the weight reading based on the calibration. This sensor measures a unique value, which is the overall weight (Hook-load) of the string including the weight of the block and Top Drive System (TDS). For all of these circumstances correct calibration is required in order to have proper reading for this drilling parameter.[22]

RPM: This parameter stands for "revolution per minute". It represents the rotational speed of the drill string. With the invention of TDS; the reading is directly linked to the electronics of the unit itself. It is considered that the measurements for this parameter are accurate as long as the acquisition system set-up has been thoroughly made up.[22]

Figure 11: Frequency sensor of RPM [23]

ROP: This parameter is the most important parameter, since all of the calculations in this study are based on estimations of Rate of Penetration (ROP). It is measured through the relative change of the position of the bit in time. Accurate calibrations are very important in order to have a representative ROP parameter**.**[22]

Torque: This parameter is the torque of the drill string while it is rotating. It is measured by means of TDS systems. Previously the readings for this parameter were relative. This parameter is going to be significantly important for inclined and highly deviated wellbores, which is also related with the wellbore cleaning issues**.**[22]

Figure 12: Frequency sensor of torque [23]

Fluid Properties: Rheological properties and the density of the drilling fluids are also among the very important parameters to be recorded for optimization purposes. Usually, the drilling fluid density is measured through calibrated MW sensors. Rheological properties on the other hand are still measured manually. Recent developments in regards to real-time pipe viscometers dictate alternative solutions. Continuous real-time viscometer probes placed on the flow line (which are reportedly under development) could facilitate data acquisition over the rheological properties of drilling fluids in real-time.[22]

LWD: Logging While Drilling, information related parameters could be captured during drilling and be used in the optimization process.[22]

Pump Parameters: The pump parameters are composed of the liner size in use, pump strokes, and the pump pressure. In case there are two pumps working simultaneously all of the data for two of the pumps should be acquired. With the electric pumps the stroke is transmitted in the same way as RPM. The pressure at the pump in case of having been acquired could be compared with the reliability of the standpipe pressure. Pump pressure should always be greater than the standpipe pressure. Use of flow meters could also be adapted for accurate flow rate measurements.[22]

Mud flow rate: In order to lubricate and cool down the bit under drilling process, a mixture of additives mixed in water or oil, which, respectively, is called water-based and oil-based drilling mud, is pumped through the drill pipe down to the bit. Drilling mud also cleans up the bottomhole by transporting the cuttings up to the surface. It also helps penetration rate as it passes bit nozzles and penetrates the rock as a water jet system. Mud flow rate is often expressed in gallons per minute (g/m) [24].

I.7 Factors Affecting Rate of Penetration:

The factors which affect rate of penetration are exceedingly numerous and perhaps important variables exist. Which are unrecognized up to this time [25].

The factors known to have an effect on rate of penetration are listed under two general classifications such as controllable and environmental. Controllable factors are the factors which can be instantly changed such as weight on bit, bit rotary speed, hydraulics.

Environmental factors on the other hand are not controllable such as formation properties, drilling fluids requirements. The reason that drilling fluid is considered tobe an environmental factor is due to the fact that a certain amount of density is required in order to obtain certain objectives such as having enough overpressure to avoid flow of formation fluids. Another important factor is the effect of the overall hydraulics to the whole drilling operation which is under the effect of many factors such as Lithology, type of the bit, down hole pressure and temperature conditions, drilling parameters and mainly the rheological properties of the drilling fluid [26].

Rate of penetration performance depends and is a function of the controllable and environmental factors. It has been observed that the drilling rate of penetration generally increases with decreased of Equivalent Circulating Density (ECD).Another important term controlling the rate of penetration is the cuttings transport. Ozbayogluet al.,(2004) [27], conducted extensive sensitivity analysis on cuttings transport for the effects of major drilling parameters, while drilling for horizontal and highly inclined wells. It was concluded that average annular fluid velocity is the dominating parameter on cuttings transport, the higher the flow rate the less the cuttings bed development.

One of the most important considerations in order to have an efficiently cutting transported hole is to take into account the factors given in table. 1[28]

Table 1: Factors for efficient hole cleaning [28]

1	Hole angle
2	Fluid Velocity
3	Fluid Properties (rheological properties and density)
4	Cuttings Size, shape, and concentration
5	Annular size
6	Rate of pipe rotation and pipe eccentricity

7	Fluid flow regime (laminar or turbulent)

Chapter II:

Artificial intelligence

II.1 History:

Artificial Intelligence (AI) has found extensive usage in simplifying complex decision-making procedures in practically every competitive market field, and oil and gas upstream industry is no exception to it. AI involves the use of sophisticated networking tools and algorithms in solving multifaceted problems in a way that imitates human intellect, with the aim of enabling computers and machines to execute tasks that could earlier be carried out only through demanding human brainstorming. Unlike other simpler computational automations, AI enables the designed tools to learn" through repeated operation, thereby continuously refining the computing capabilities as more data is fed into the system.

Over the years, AI has led to significant designing and computation optimizations in the global Petroleum Exploration and Production (E&P) industry, and its applications have only continued to grow with the advent of modern drilling and production technologies. Tools such as Artificial Neural Networks (ANN), Generic Algorithms, Support Vector Machines and Fuzzy Logic have a historic connection with the E & P industry for more than 16 years. Now, with the first application dated in 1989 for development of an intelligent reservoir simulator interface, and for well-log interpretation and drill bit diagnosis through neural networks. Devices and software with basis from the above mentioned AI tools have been proposed to abridge the technology gaps hindering automated execution and monitoring of key reservoir simulation, drilling and completion procedures including seismic pattern recognition, reservoir characterization and history matching, permeability and porosity prediction, PVT analysis, drill bits diagnosis, overtime well pressure-drop estimation, well production optimization, well performance projection, well/ field portfolio management and quick, logical decision making in critical and expensive drilling operations[29].

II.2 definition:

Artificial intelligence (AI) is a branch of computer science that in its broadest sense would mean the ability of a machine to perform or in an attempt to mimic the operations of human brains or human thought, understand and apply thinking methodologies. It is an approach of solving engineering difficulties with the ability to consider all effective parameters simultaneously. The central principles of AI include reasoning, knowledge, planning, learning, communication, perception and the ability manipulate objects.[29]

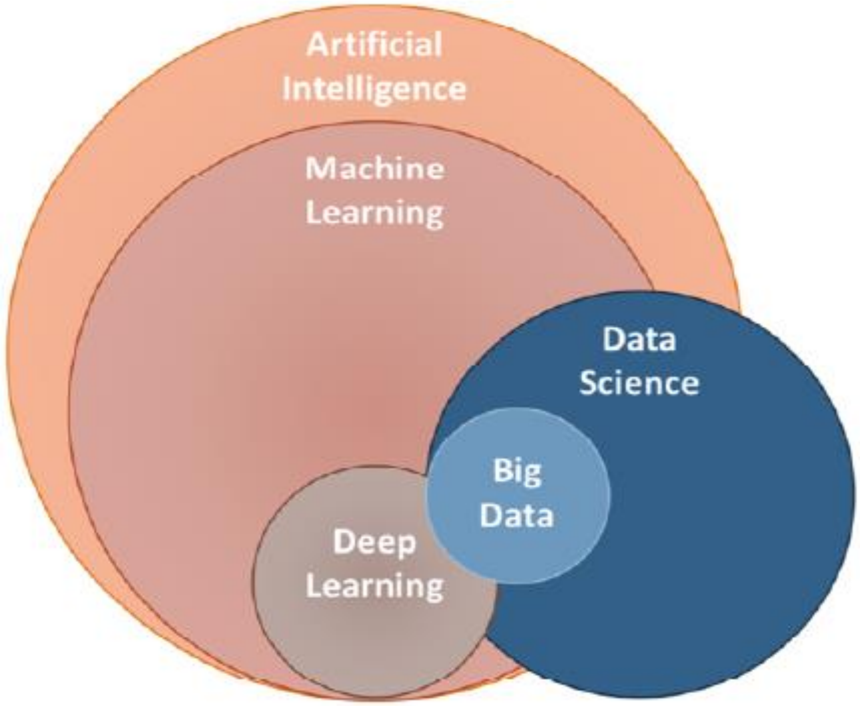


Figure 13: A clear visualization of overlapping Artificial Intelligence related terminology [30]

II.3 Artificial intelligence subfields:

II.3-1 Machine learning: Machine learning is a subset of artificial intelligence, it is the science of getting computers to act without being explicitly programmed, and is mostly just statistics. Machine learning is used to find patterns in data that you can then make predictions on. It can be subdivided into supervised learning and unsupervised learning or some mixture of both[31].

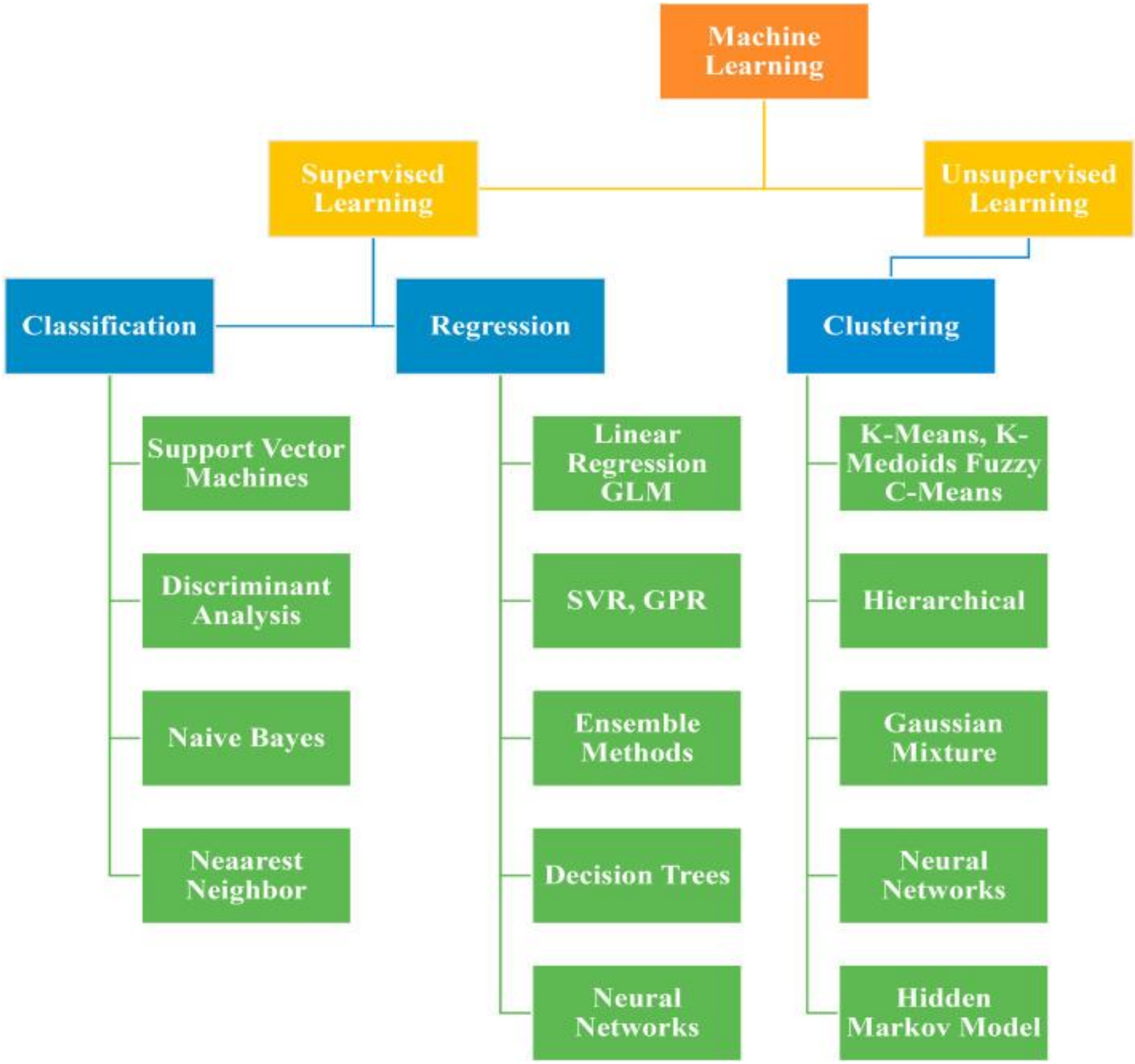


Figure 14: Machine learning classification [32]

Machine learning is a computer program said to learn from experience 'E' with respect to some class of tasks 'T' and performance measure 'P', if its performance at tasks in 'T', as measured by 'P', improves with experience 'E' [31].

II.3-2 Natural Language Processing:

Natural-language-processing software package use AI to let a user to communicate with a computer in the user's natural language. The computer can both recognize and reply to instructions given in a natural language. The penalty area of Natural Language Processing (NLP) is to plan and construct a computer system that will investigate, recognize, and produce natural human-languages [31].

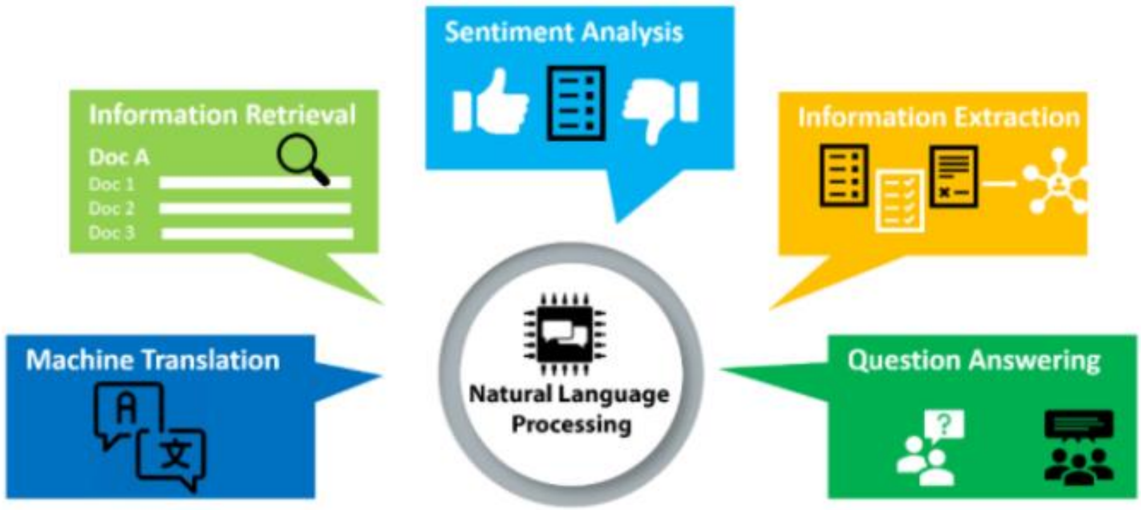


Figure 15: Applications of NLP [33]

II.3-3 Robotics:

Robotics is the area of Artificial Intelligence technology most pretty to the community. In fact, robotics could be the area where AI can be supreme beneficial to mankind. It includes mechanical, generally computer-controlled devices to accomplish tasks that necessitate extreme accuracy or monotonous or dangerous work by people [31].

II.3-4 Fuzzy Logic:

is a technique that represents and modifies uncertain information by measuring the degree to which the hypothesis is correct. Fuzzy logic is also used for reasoning about naturally uncertain concepts. Fuzzy logic is convenient and flexible to implement machine learning techniques and assist in imitating human thought logically [31].

It is simply the generalization of the standard logic where a concept exhibits a degree of truth between 0.0 to 1.0. If the concept is completely true, standard logic is 1.0 and 0.0 for the completely false concept. But in fuzzy logic, there is also an intermediate value too which is partially true and partially false.

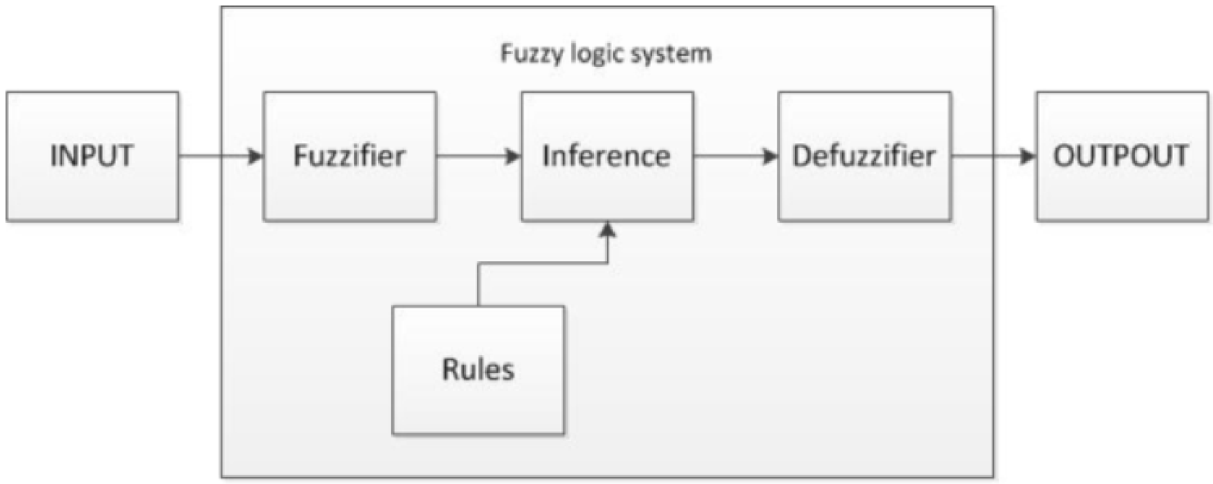


Figure 16: Fuzzy logic system [34]

II.3-5 Expert system:

refers to a computer system that mimics the decision-making intelligence of a human expert. It conducts this by deriving knowledge from its knowledge base by implementing reasoning and insights rules in terms with the user queries.

The effectiveness of the expert system completely relies on the expert's knowledge accumulated in a knowledge base. The more the information collected in it, the more the system enhances its efficiency. For example, the expert system provides suggestions for spelling and errors in Google Search Engine [31].

II.4 Machine learning for oil field:

ML techniques helped in reducing the risks associated with drilling wells. The original application was developed in 1989, it utilized neural networks for the progress of drill bit diagnostics, well-log interpretation, monitoring of important drilling and completion operations and reservoir simulation including seismic pattern recognition, reservoir characterization and history matching, permeability and porosity prediction, PVT analysis, well pressure-drop assessment, production optimization, well performance prognosis, well / field portfolio management. Well construction is carried out on the base of project documentation, and the accuracy of project data has a considerable influence on the technical and economic success of drilling, along with its environmental safety and prevention of accidents. In many situations, drilling design is based on past data and experience. However, this seems to be ineffective with depleted reservoirs and complex geology and that is where machine learning is utilized. By setting up quick responses of protection procedures and failure identification or self-learning on well design tools, the chances of drilling success could increase achieving lower cost per foot. Developed an ML algorithm to calculate alarm thresholds, adapted in real time, to find anomalies in mud volume and flow rate data throughout the entire drilling process. This methodology contributed to early lost circulation detection and helped minimized cumbersome false alarms. NN replicated directional drillers' expertise to minimize tortuosity and deviation from planned wellbore trajectory while maximizing ROP with less than 3% error. System automation using data from sensors and control algorithms aims at reducing human error, promoting safety of personnel and improving productivity and efficiency, and drilling parameter optimization. If a parameter diverts from the specified range, the system corrects set points to bring he parameter back into range. One example of such technology was the improvement of rate of penetration (ROP), established a closed loop system which checked real-time drilling operations for ROP maximization through weight on bit (WOB) and RPM set point adjustment for ROP optimization. Several ML algorithms can be used to obtain drilling parameters' trend analysis, detect abnormal events throughout different drilling phases, and accordingly suggest corrective measures.[35]

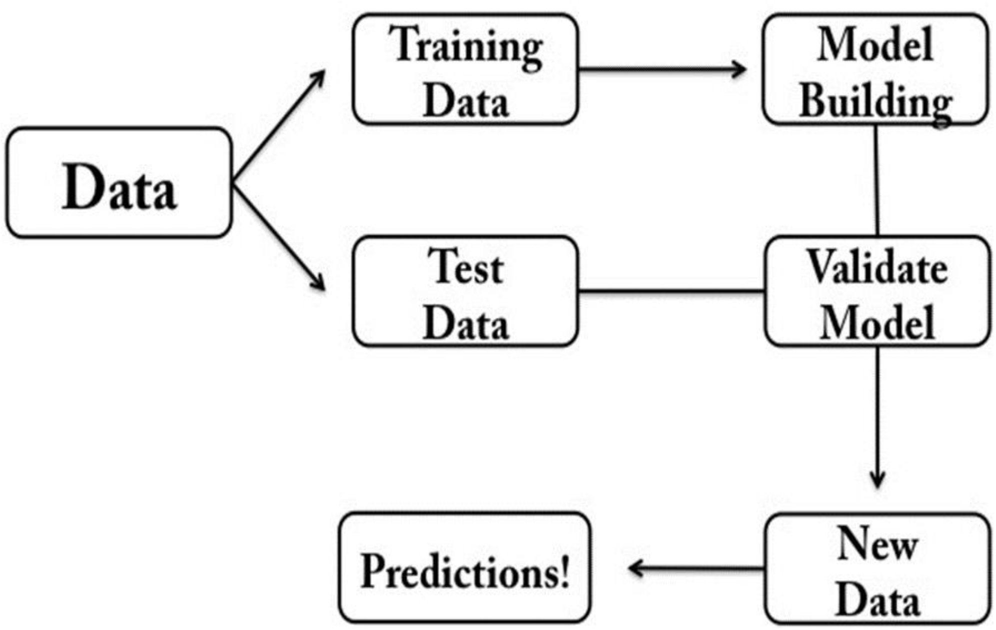


Figure 17: Machine learning process [36]

Chapter III:

Data mining

III.1 Definition:

is a component of a wider process called "knowledge discovery from data base". It involves scientists and statisticians, as well as those working in other fields such as machine learning, artificial intelligence, information retrieval and pattern recognition.

III.2 CRISP-DM:

Any good project starts with a deep understanding of the problem initially and followed by providing an efficient and effective solution to resolve the business issues. In typical analytics projects which involve multiple steps like data cleaning, preparation, modeling, and model evaluation, a framework for recording experience and is needed to allow projects to be replicated. CRISP-DM stands for **Cr**oss **I**ndustry **S**tandard **P**rocess for **D**ata **M**ining and is an open-source and widely used methodology created to shape Data Mining projects. This comprehensive methodology provides anyone with a complete blueprint for conducting a data mining project. The process breaks down the lifecycle of a data mining project into six phases to encourage best practices and help to obtain better results overall.[37] (**Figure 36**)

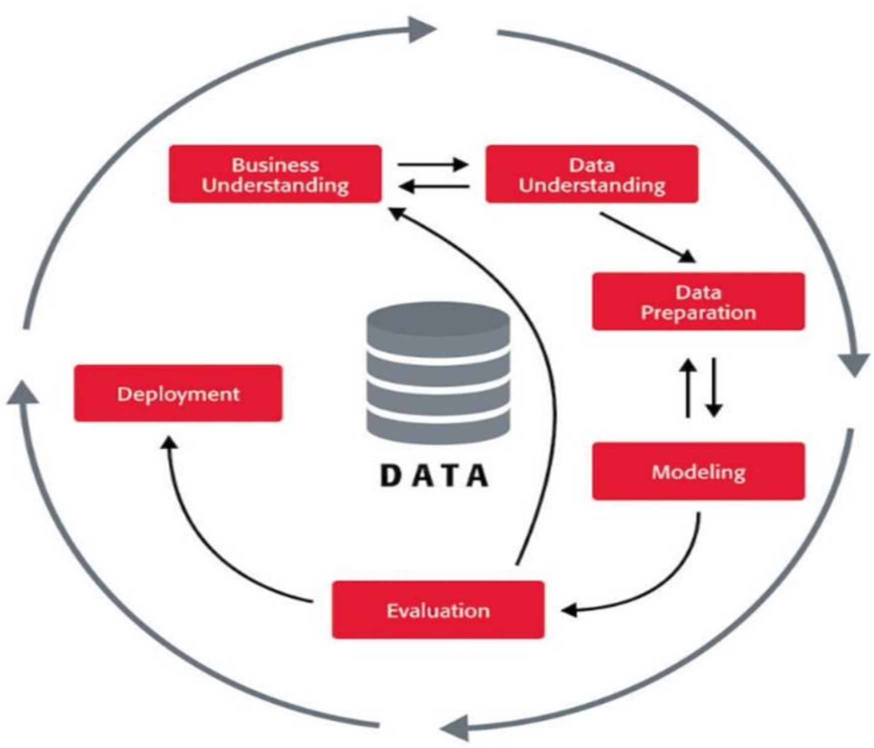


Figure 18: CRISP-DM overview and how a data mining project breaks into six phases to obtain better results [37]

The six phases of CRISP-DM include:

1) Business Understanding: In this step, the goals of the businesses are set and the important factors that will help in achieving the goal are discovered.

2) Data Understanding: This step will collect the whole data and populate the data in the tool (if using any tool). The data is listed with its data source, location, how it is acquired and if any issue encountered. Data is visualized and queried to check its completeness.

3) Data Preparation: This step involves selecting the appropriate data, cleaning, constructing attributes from data, integrating data from multiple databases.

4) Modeling: Selection of the data mining technique such as decision-tree, generate test design for evaluating the selected model, building models from the dataset and assessing the built model with experts to discuss the result is done in this step.

5) Evaluation: This step will determine the degree to which the resulting model meets the business requirements. Evaluation can be done by testing the model on real applications. The model is reviewed for any mistakes or steps that should be repeated.

6) Deployment: In this step a deployment plan is made, strategy to monitor and maintain the data mining model results to check for its usefulness is formed, final reports are made and review of the whole process is done to check any mistake and see if any step is repeated.

CRISP-DM model is an idealized sequence of events that creates a long-term strategy by structuring a basic and simple but still “good enough” model during the first iteration and improved the model in further iterations. Following CRISP-DM guidelines, a leading approach for managing data mining and predictive analytic for big data [37].

III.3 Data-mining process:

Before a data set can be mined, it first has to be “cleaned”. This cleaning process removes errors, ensures consistency and takes missing values into account. Next, computer algorithms are used to “mine” the clean data looking for unusual patterns. Finally, the patterns are interpreted to produce new knowledge.

The data mining process is divided into two parts. Data Preprocessing and Data Mining. Data Preprocessing involves data cleaning, data integration, data reduction, and data transformation. The data mining part performs data mining, pattern evaluation and knowledge representation of data.[38]

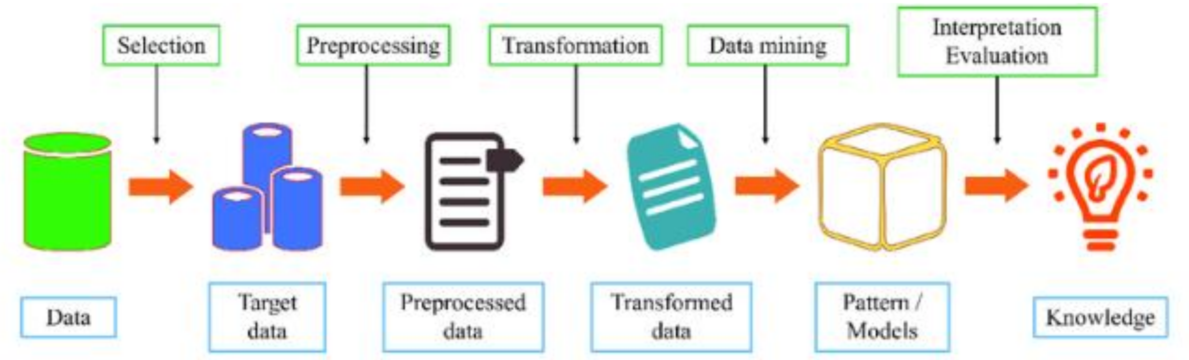


Figure 19: Data mining process [39].

Data Preprocessing in Data Mining

Data preprocessing is a data mining technique which is used to transform the raw data in a useful and efficient format.

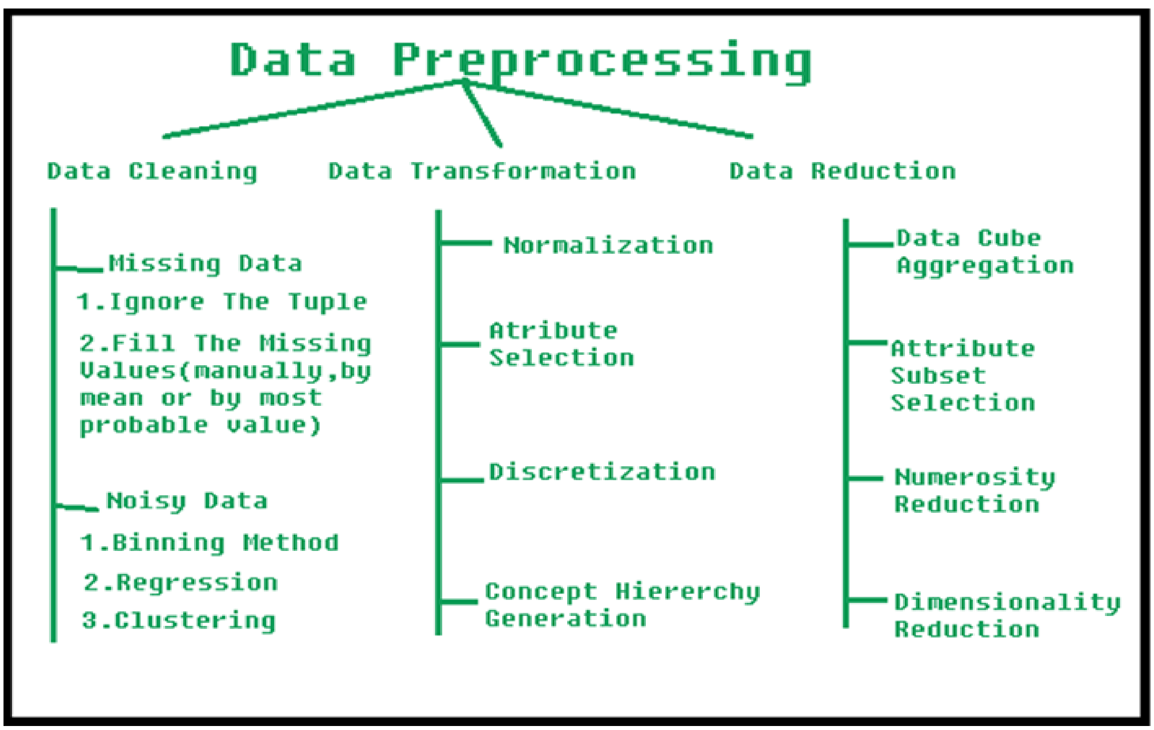


Figure 20: Data Preprocessing in Data Mining [40]

Data Cleaning is the first step in data mining. It holds importance as dirty data if used directly in mining can cause confusion in procedures and produce inaccurate results. Basically, this step involves the removal of noisy or incomplete data from the collection. Many methods that generally clean data by itself are available but they are not robust.[40]

This step carries out the routine cleaning work by:

- ✓ **Fill The Missing Data**

```
DataFrame.fillna(value=None, method=None, axis=None, inplace=False, li
mit=None, downcast=None)
```

- ✓ **Remove The Noisy Data**: Random error is called noisy data

```
DataFrame.filter(items=None, like=None, regex=None, axis=None)
```

Data integration: When multiple heterogeneous data sources such as databases, data cubes or files are combined for analysis, this process is called data integration. This can help in improving the accuracy and speed of the data mining process [40].

Different databases have different naming conventions of variables, by causing redundancies in the databases. Additional Data Cleaning can be performed to remove the redundancies and inconsistencies from the data integration without affecting the reliability of data.[40]

Data Reduction: this is helpful to reduce the data dimensionality as some of the parameters in a dataset are not going to make any contribution to the final result. Moreover, working with extensive data causes several problems such as complex and time-consuming modeling [40].

Some strategies of data reduction are:

- ✓ **Dimensionality Reduction:** Reducing the number of attributes in the dataset.
- ✓ **Numerosity Reduction:** Replacing the original data volume by smaller forms of data representation.
- ✓ **Data Compression:** Compressed representation of the original data.

Data Transformation In this process, data is transformed into a form suitable for the data mining process. Data is consolidated so that the mining process is more efficient and the patterns are easier to understand. Data Transformation involves Data Mapping and code generation process [40].

Strategies for data transformation are:

- ✓ **Smoothing:** Removing noise from data using clustering, regression techniques, etc.
- ✓ **Aggregation:** Summary operations are applied to data.
- ✓ **Normalization:** Scaling of data to fall within a smaller range.
- ✓ **Discretization:** Raw values of numeric data are replaced by intervals.

Data Mining In this step, intelligent patterns are applied to extract the data patterns. The data is represented in the form of patterns and models are structured using classification and clustering techniques [40].

Pattern Evaluation This step involves identifying interesting patterns representing the knowledge based on interestingness measures. Data summarization and visualization methods are used to make the data understandable by the user [40].

Knowledge Representation is a step where data visualization and knowledge representation tools are used to represent the mined data. Data is visualized in the form of reports, tables [40]

Chapter IV:

Matriels and Methods

IV.1 Dataset:

A section of 1000 m (16" phase) of drilling datasetwas selected from a well in Hassi Messouad field to perform this study. The used raw dateset is shown in table(V.7) and cleaned data.

This part of the well was selected because of the homogeneous of its lithological characteristics and to avoid the effect of the lithology on the chosen model.

The description of this section are presented in the paragraph "Description of the selected phase."

The name and the coordinnets of the well was deleted from the given data for confidential purpose. The Well data consist of real-time drilling parameters such as well depth, rate of penetration, weight on bit, flow rate, rotation per minute, torque, bit diameter, stand pipe pressure, etc

In this study there are some unknown parameters co-efficient which must be determined based on past drilling data obtained from a field in order to determine theseco-efficient, a linear regression technique will be applied which as follows.

$$Y= b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \ldots\ldots + b_nX_n \quad \text{……………………………………………………………………..IV.1}$$

Where Y is the dependent variable, b_0 is the intercept term and the regression co-efficient b_1,b_2,b_3, ------bn are the analogues of the shape of linear regression. From the above equation Y is the ROP relevant drilling parameters will make up the regression variable [X_1&X_n]. b_0 to bn Co-efficient will be determined by machine learning using a programming languagesoftware called python and its packages. This python software will perform the regression analysis after all the relevant drilling parameters has been uploaded into it and then run. The analysis will then provide an output computed data model.

IV.2 Python

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.

Python is Interpreted − Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.

Python is Interactive − You can actually sit at a Python prompt and interact with the interpreter directly to write your programs.

Python is Object-Oriented − Python supports Object-Oriented style or technique of programming that encapsulates code within objects.

Python is a Beginner's Language and a great language for the beginner-level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games.

Table 2:Python description and characteristics

First version date	February 20, 1991
Paradigms	Object, imperative and functional
Author	Guido van Rossum
Developers	Python Software Foundation
Last version	3.10.5 (6 June 2022)
Version in development	3.11.0b3 (June 1, 2022)
Typing	Strong, dynamic, duck typing
Influenced by	ABC, C, Eiffel, ICON, Modula-3, Java, Perl, Smalltalk, Tcl
Influenced	Ruby, Groovy, Boo, Julia
Implementations	CPython, Jython, IronPython, PyPy
Written in	C for CPython, Java for Jython, C# for IronPython and in Python itself for PyPy
Operating system	Multiplatform
Licence	Free license -Python Software Foundation License
Website	www.python.org
File extension	py, pyc, pyd, pyo, pyw, pyz et pyi

IV.3 Jupyter Notebook

JupyterLab is the latest web-based interactive development environment for notebooks, code, and data. Its flexible interface allows users to configure and arrange workflows in data science, scientific computing, computational journalism, and machine learning. A modular design invites extensions to expand and enrich functionality [41].

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.

Jupyter notebook is running using your browser, it could run locally on your machine as a local server or remotely on a server. The reason it is called notebook is because it can contain live code, rich text elements such as equations, links, images, tables, and so on. Therefore, you could have a very nice notebook to describe your idea and the live code all in one document. Thus Jupyter notebook becomes really popular way to test ideas, writing blogs, papers and even books, for example, this book is written entirely within Jupyter notebook. Of course, it has many other advantages, and we will only cover the basics of the Jupyter notebook to get you started [42].

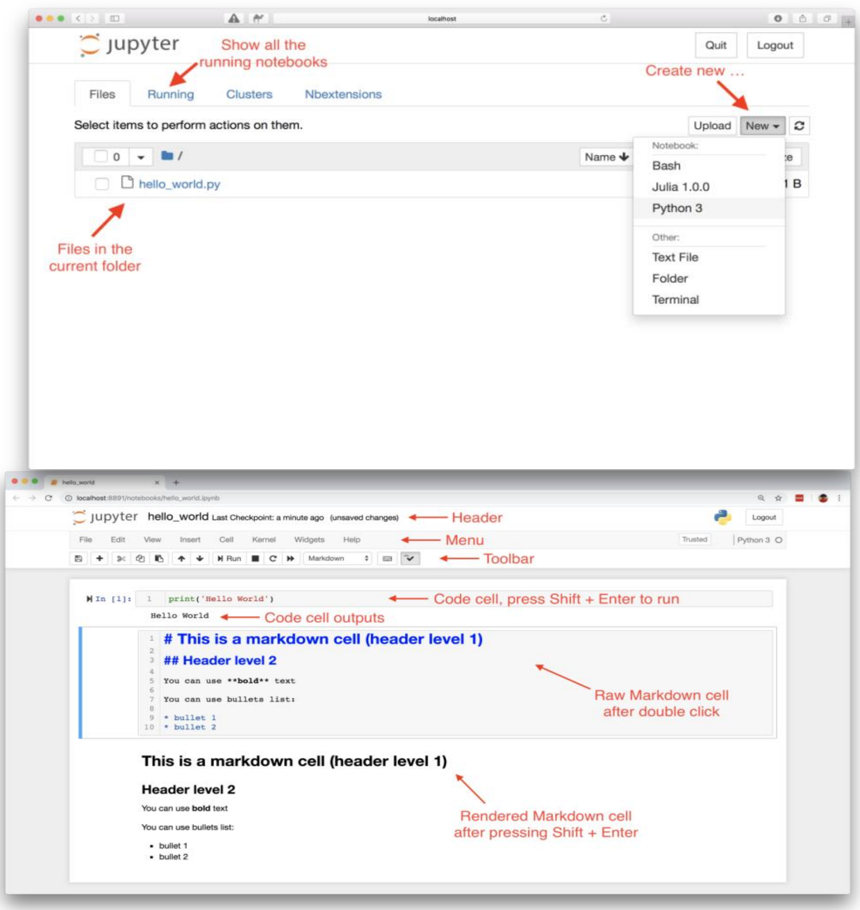


Figure 21:Jupyter Notebook,the Classic Notebook Interface [41]

IV.4 Python libraries and packages:

Python Libraries are a set of useful functions that eliminate the need for writing codes from scratch.There are over 137,000 python libraries present today.

Python libraries play a vital role in developing machine learning, data science, data visualization, image and data manipulation applications, and more. Let us start with a brief introduction to Python Programming Language and then directly dive into the most popular Python libraries.

The probability that you must have heard of 'Python' is outright. Guido Van Rossum's brainchild – Python, which dates back to the '80s has become an avid game changer. It is one of the most popular coding languages today and is widely used for a gamut of applications. In this article, we have listed 34 Open Source Python Libraries you should know about.

IV.5 Model selection:

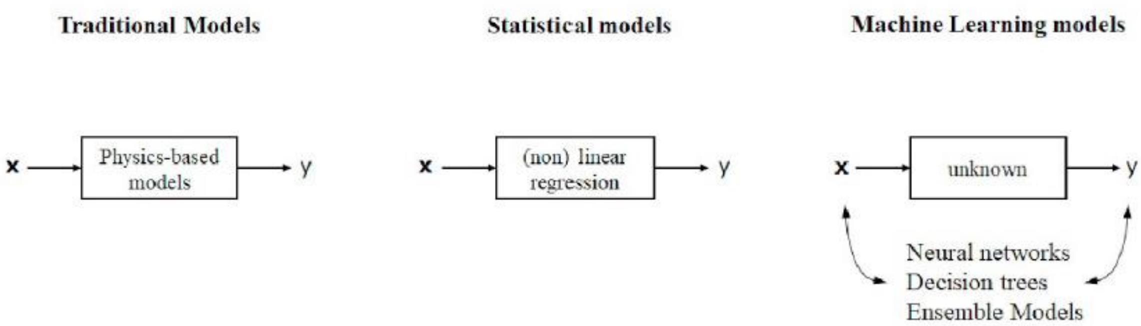


Figure 22: Approaches for ROP modeling Breiman (2001b) [44].

To expect the rate of penetration (ROP), it is required to have a model able to evaluate how the drilling parametres affect ROP. In a general structure, obtaining a ROP model can be thought of as a regression problem. Following the notation adopted by Friedman (2006), the prediction of ROP can be formulated as

yˆ = f (x,) (1) : where yˆ represents the estimation for the output variable (i.e. ROP), x represents the vector of inputs or predictors (e.g. drilling parameters, such as drill bit type, torque, WOB, mud flow), f ( ) the regression function, and the regression function's parameter set.

1-**machine learning algorithms**:linear regression

-Decission tree - KNN(k-Nearest Neighbors)

-RF (Random Forest) -SVM(supportvector machine)

2-linear regression ⟹ suprevised learning ⟹ values prediction

Regression analysis is used to estimate the relationships among one dependent and two or more independent variables .This method of data analysis is useful when examining a quantitative variable in relation to other factors. The Multivariate analysis describes an observation factor by having several variables, taking into consideration all changes of properties that may happen simultaneously.[45]

2-1-A simple linear regression model is a mathematical equation that allows us to predict a response for a given predictor value [46].

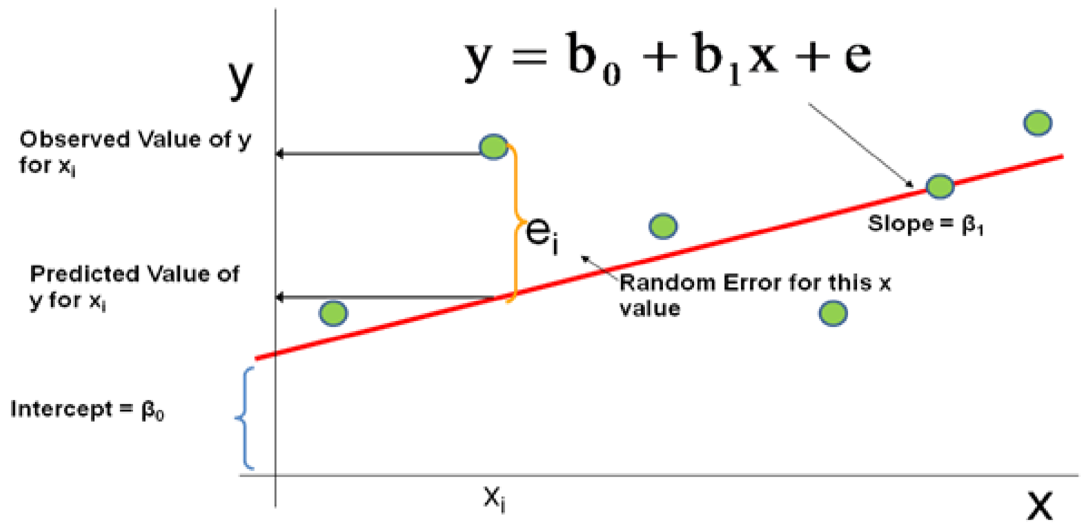


Figure 23:simple linear regression[47]

Our model will take the form of $Y = b_0 + b_1x+e$ where b_0 is the y-intercept, b_1 is the slope or regeression coefficient, x is the predictor variable, and $Y=y_{predict}$ an estimate of the mean value of the response variable for any value of the predictor variable.

2-2-the multiple regression equation of (Y) factor on variables (X_1, X_2, …, Xn) is given by:

$$Y= b_0 + b_1X_1 + b_2X_2 + b_3X_3 + ...+ b_nX_n \qquad \text{IV.2}$$

(b_1, b_2, b_3…, b_n) are the analogues to the slope in linear regression equation.

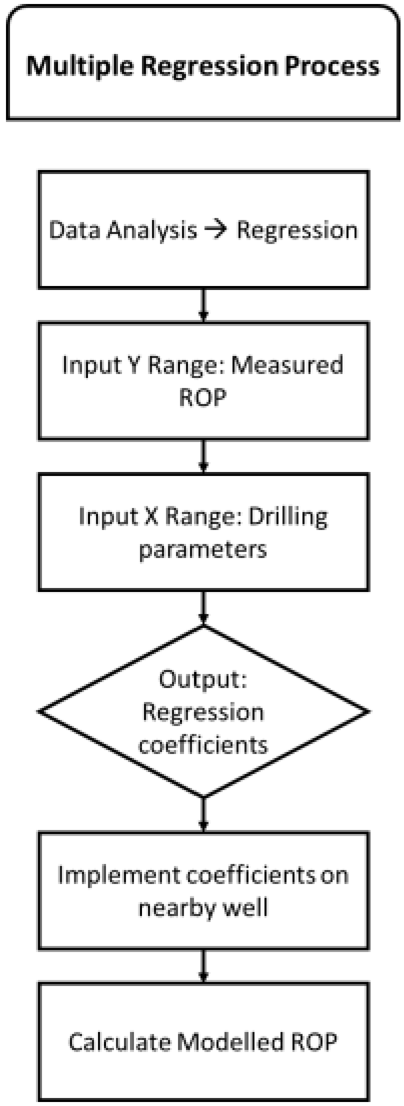


Figure 24:Least squares process flowchart [45]

3-model=LinearRegression()

```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn import linear_model
import numpy as np
```

4-1-estimitation selection-hyper parameter training

model=linear regresion

4-2-Model training

Model.fit(x,y)

```
lm=linear_model.LinearRegression()
model=lm.fit(x, y)
```

4-3-Evaluation (model evaluation)

Model.score(x,y)

```
#evaluate the model
model.score(x, y)
```

4-4-Model implementation

Chapter V:

Results and discussion

DATA EXPLORATION

V.1 Description of the selected phase (16''):

Drilling data available for the study was acquired from a field located in Hassi Messaoud.Is formed of different formation classifications.The mud weight used for drilling this phase is density oil (1.18 to 1.25 SG).The objective of this intermediate phase is to tube in 13"3/8 the formations of the Cretaceous and part of the Jurassic (Clay and Lagoon).

Table 3: gives the detailed table for the formations Lithology description.

Table 3: formations Lithology description

Age	Depth	DESCRIPTION
Mio-Pliocene	11-235m	-Sands. -Hard siliceous with limestone intercalation.
Eocene	235-338m	-Sands. -hard limestone. -micro crystalline.
Senonien Carbonate	338-437m	-Dolomitic limestone -Anhydrite -Dolomite
Senonien Anhydritique	437-642 m	-Anhydrite - Dolomite -Clay

V.1 .1 Specifications of the 16'' Phase:

- Realization of the Section: 1784 m
- drilling casing 18 5/8" with PDC bit.
- EI oil based mud d=1.25sg (mud density 1.25 before the top of the Turonian)
- Drilling from 506m to 2290m (110m in the Dogger Lagunaire)
- Descent of a 13" 3/8 column.
- Cementation in a single phase.

We can expect a slowdown in advancement in the Aptian because of the dolomite, and draws in the salt zone, especially during possible maneuvers. It is a long phase usually 1880 meters and drilled ideally with a single PDC tool.

The phase specification is the presence of aquifers and the Senonian salt formation

Turonian aquifer: Salt water, its salinity in Hassi Messaoud varies from 164 to 240 g/l, it is not drinkable. The porosity of the aquifer is about 24%, the water is estimated at 2586,7. 10^9 m^3 and the pore pressure control density is about 1.03.

Albian aquifer: Fresh water, the salinity varies from 0.24 to 9.5 g/l, it is moderately drinkable. The porosity of the aquifer is estimated at 20.99%, the wateris estimated at 5,508.10^{13} m^3.The Albiandensity at Hassi Messaoud is estimated at around 123 bars (deq=1.17).

The following table shows the percentage of clay in each layer.

Table 4: The percentage of clay in each layer

Layer	Clay %
Sen Salifere	/
Turonian	26.60
Cenomanian	10.40
Albian	4.00
Aptian	14.40
Neocomian	87.80
Malm	95.00
Clay Dogg	55.00
Lagonn Dogg	5.00

V.1 .2 Objectives Of Phase 16" :

- The main objectives of this phase are the following :
- Drilling the section in one run.
- Zero lost time accident (LTA).
- Lay and cement the 13"3/8 110mDV in the lagoon dogger.
- Ensure a good insulation of the Albian and Barremian terrains.
- No harm to the environment.

V.1 .3 drilling parameter :

WOB = 5 -20 tonnes \ RPM = 80 - 150 rpm

Q = 2800 - 3200 lpm \ d= 1.18 – 1.25 sg

V.1 .4configuration of BHA

Table 5: BHA description.

Materials	Characteristics			
16" PDC Bit	1	0.40 m	0.00	16,00
Near bit stabilizer 16"	1	2.00 m	3.00	15,88
9"1/2 Short Drill Collar	1	3.62 m	3.00	9,50
String Stabilizer 16"	1	2.23 m	3.00	15,88
9" 1/2 Drill Collar	1	9.16 m	3.00	9,50
String Stabilizer /16"	1	2.23 m	3.00	15,88
9" 1/2 Drill Collar	2	18.12 m	3.00	9,50
XOS	1	1.12 m	3.00	9,50
8" Drill Collar	9	81.79 m	2.81	8,00
Hydraulic Jar 8''	1	9.90 m	2.81	8,00
8" Drill Collar	2	18.17 m	2.81	8,00
XOS	1	1.09 m	2.81	8,00
5 1/2" HWDP	12	110.84 m	3.63	5,50
Total Length		260.67 m	\	

V.1 .5Mud parameter :

Table 6:Mud parameter

System		
Sepecific gravity	/	1.20 -1.25
Plastic viscosity	120 °F/Cps	As low as possible
Yield point	lb/100 ft2	18-24
Gels strength 10''	lb/100 ft2	/
Filtrat HP/HT	250 °F/500psi	<10
Electrical stability	Volt	>600
Oil/water ratio	/	70/30 - 85/15
Sand	% en vol	% en vol
Low gravity solids	% en vol	<5

V.2Data Analysis and Interpreting: In this party,we present to interpreting a curves analysisof drilling parameters .In this thesis, all the algorithms and analyses were developed in Jupyter Notebookwhich is a free, open-source, interactive web tool. Jupyter is an easy-to-use, interactive data science environment that is the best coding language for data mining andanalysis. Jupyter is a great interface to the Python programming language that containsmany powerfullibraries, ranging from basic statistics to complex machine learningalgorithms. Some of the commonly used libraries in this work are asfollows:

- Pandas, Numpy: These are grouped as basic libraries which are used for underlying data analys
- scikit-learn is used to build ML models:
  - sklearn.metrics: used for evaluation.
  - sklearn.preprocessing: used for scaling.
- matplotlib: is known as Python’s most powerful visualization librarie.

Table 7: Overview in an analysis parameter

```
Data columns (total 11 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   Date                          17279 non-null  datetime64[ns]
 1   Depth_TVD_m                   17279 non-null  float64
 2   Bit_Position_m                17279 non-null  float64
 3   ROP_AVG_h/m                   17279 non-null  float64
 4   ROP_AVG_m/s                   12025 non-null  float64
 5   WOB_Avg_t_metric              17279 non-null  int64
 6   Torque_Avg_Ibf.ft             17279 non-null  int64
 7   RPM_Avg_rpm                   17279 non-null  int64
 8   Pressure_SPP_ Avg_psi         17279 non-null  int64
 9   Flow_In_Pumps_(l/m)           17279 non-null  int64
 10  Mud_Temperature_OUT_Avg_(°C)  17279 non-null  float64
dtypes: datetime64[ns](1), float64(5), int64(5)
```

Table 8: Data description

	Depth_TVD_m	Bit_Position_m	ROP_AVG_h/m	ROP_AVG_m/s	WOB_Avg_t_metric	Torque_Avg_lbf.ft	RPM_Avg_rpm	Pressure_SPP_ Avg_psi	Flow_In_Pumps_(l/m)	Mud_Temperature_OUT_Avg_(°C)
count	17279.000000	17279.000000	17279.000000	12025.000000	17279.000000	17279.000000	17279.000000	17279.000000	17279.000000	17279.000000
mean	742.364989	736.857052	15.033781	0.261791	4.111928	5672.350715	92.544650	1277.621679	2478.579374	46.104786
std	106.911936	107.304682	19.330592	0.328128	5.013014	5449.587696	50.939456	530.078896	879.164292	12.796831
min	553.260000	547.550000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	20.800000
25%	657.495000	641.140000	0.000000	0.000000	0.000000	825.000000	52.000000	1208.000000	2518.000000	34.800000
50%	748.670000	740.430000	0.000000	0.000000	0.000000	2407.000000	75.000000	1543.000000	2911.000000	46.200000
75%	834.670000	833.980000	29.700000	0.526667	8.000000	10520.000000	144.000000	1609.000000	2931.000000	58.800000
max	920.590000	920.680000	103.500000	1.725000	17.000000	39827.000000	150.000000	1710.000000	7343.000000	62.200000

Table 8 gives a summary of ready-to-use data

V.2-1 Variations of parameters in terms of depth:

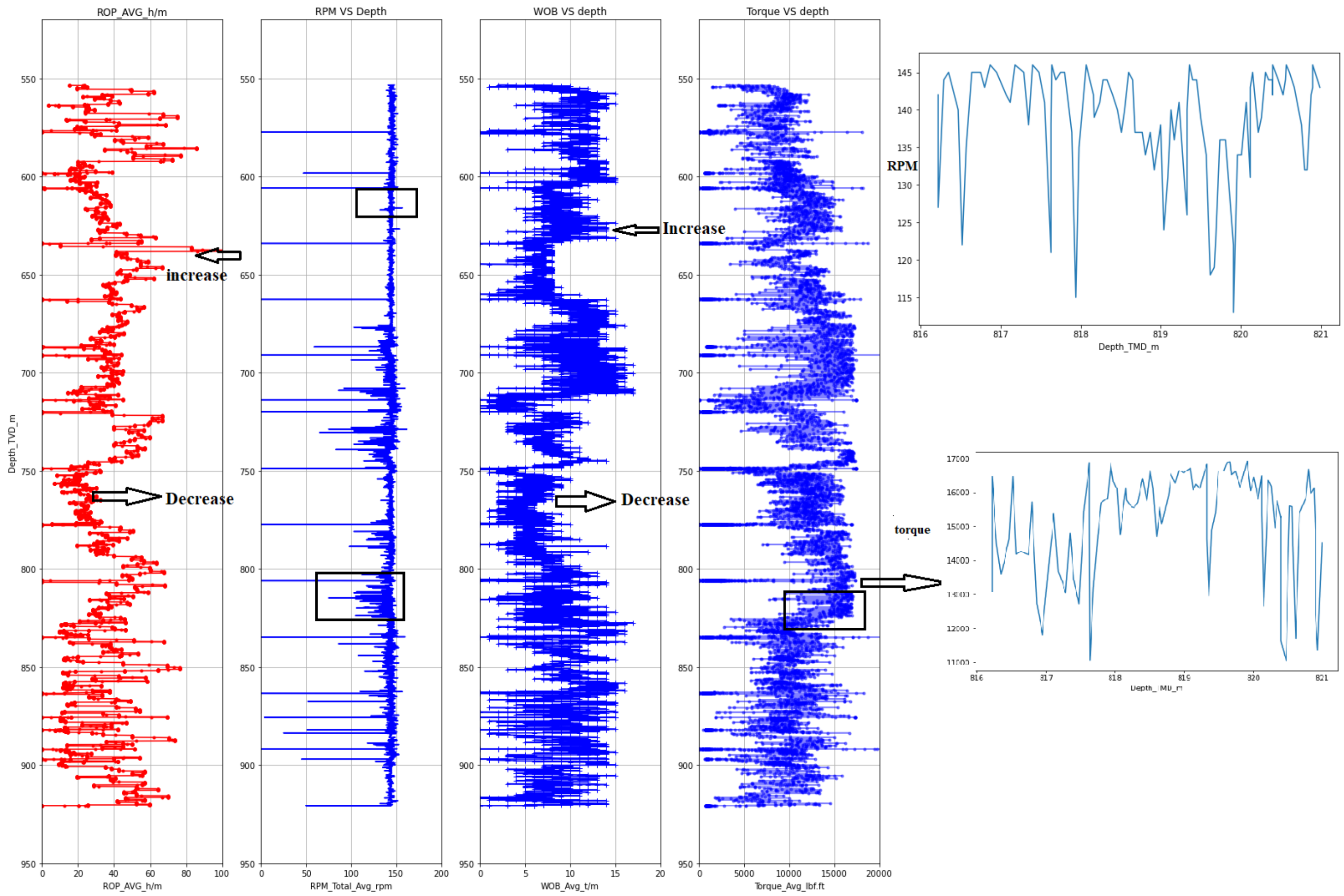


Figure 25:Drilling parameter vs depth

the curve shows drilling parameter vs depth which have a similar variation in rate of penetration and weight on bit ,in 760 m both decrease this confirms the relation between them (Figure V-29)

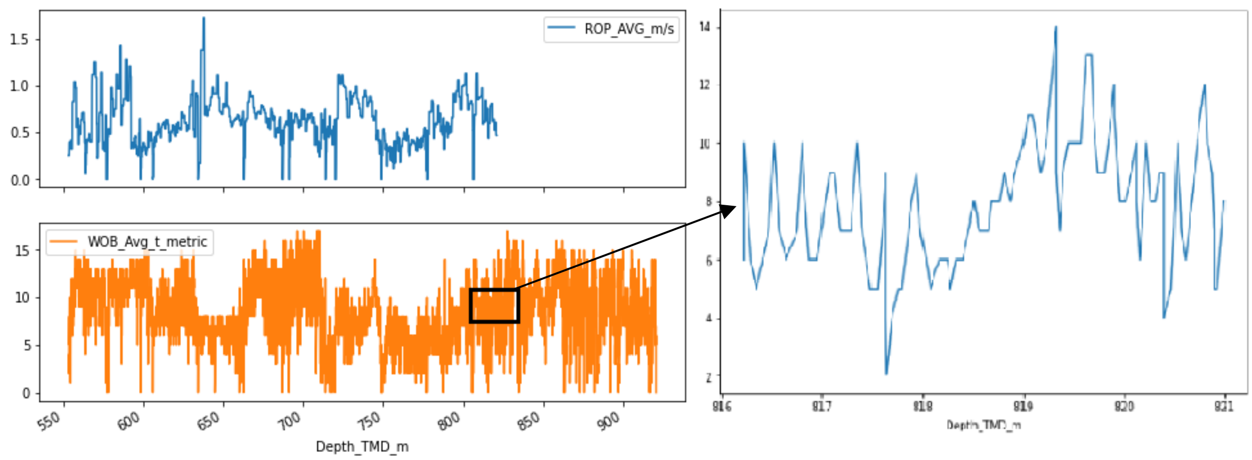


Figure 26: Correlation between ROP and WOB

One of the best ways to inspect data is to visualize it. One way to do this is by using a scatter plot. A scatter plot of the data puts one feature along the x-axis and another along the y-axis, and draws a dot for each data point. Unfortunately, computer screens have only two dimensions, which allows us to plot only two (or maybe three) features at a time. It is difficult to plot datasets with more than three features this way.Figure 27

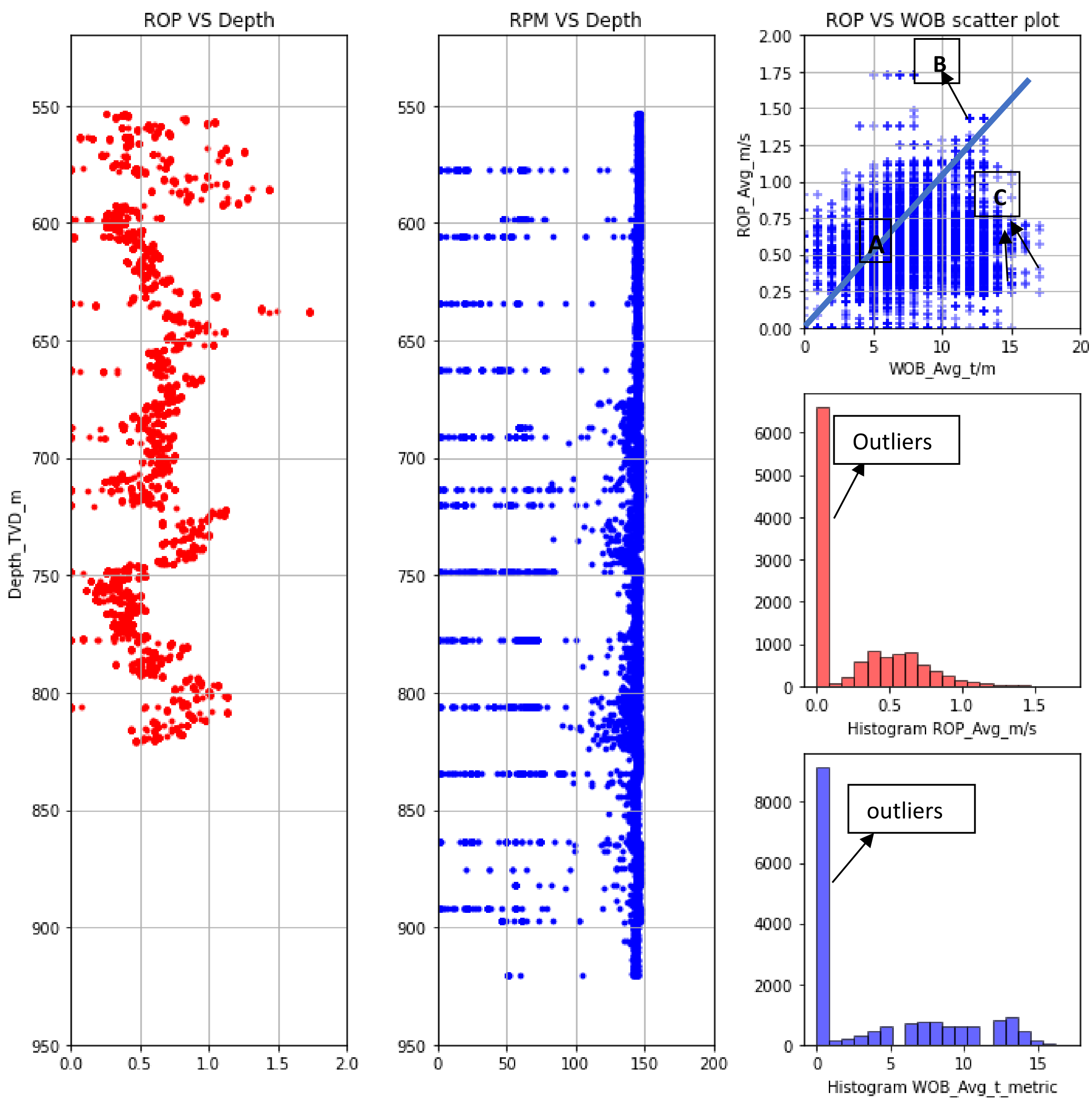
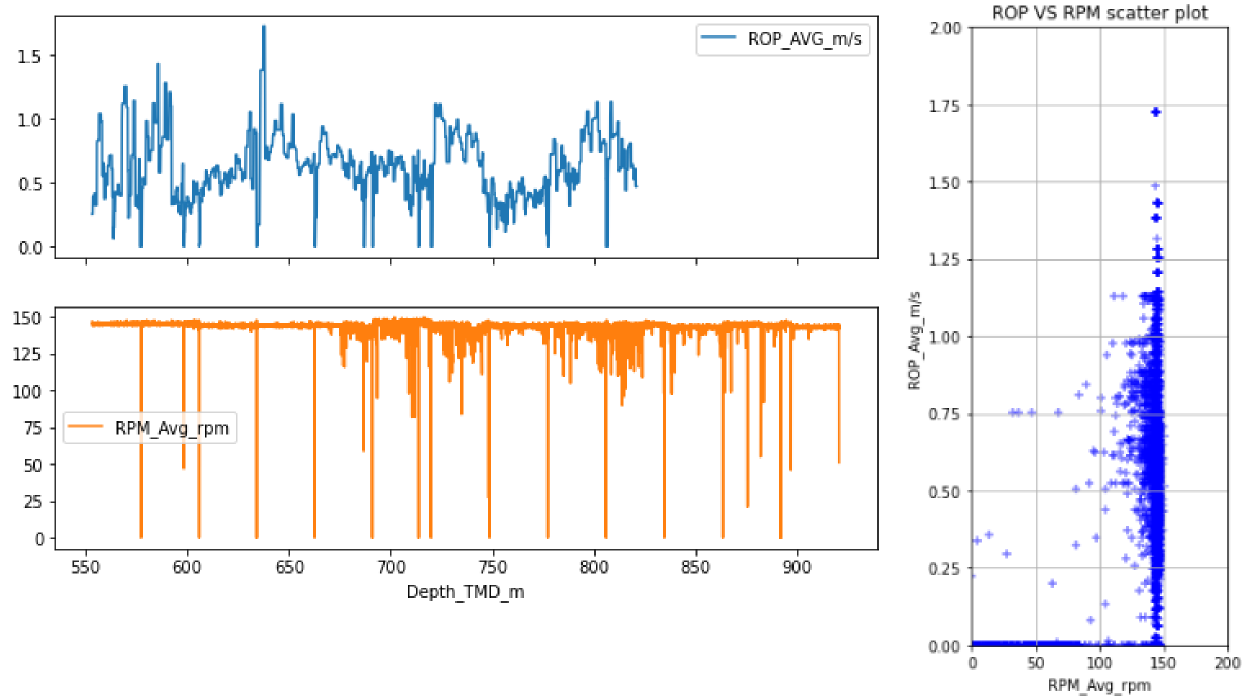


Figure 27: Scatterplot and histogram of drilling parameter

In scater plot of ROP vs WOB, we showspenetration rate increases gradually and linearly with increasing values of bit weight for low values of bit weight (point A). A linear curve is again observed at higher bit weights (pointB), representing increased drilling efficiency. Beyond Point B, subsequent increases in bit weight cause only slight improvements in penetration rate. In some cases, a decrease in penetration rate is observed at extremely high values of bit weight (point C).The poor response of penetration rate at high WOB values is usually attributed to less-efficient hole cleaning because of a higher rate of cuttings generation

The histogram of rate of penetration and weight on bit are changing with the normal average.

Figure 28: Relation between ROP and RPM



The penetration rate (ROP) usually increases linearly with increasing value of rotary speed (RPM) for low values of rotary speed. At higher values, the rate of increase diminishes. The point where we loose linearity is called the Foundering point. The phenomenon is essentially due to less efficient bottomhole cleaning and is also dependent to drilling fluid parameters (density).

V.2-2 Variations of parameters versus time :

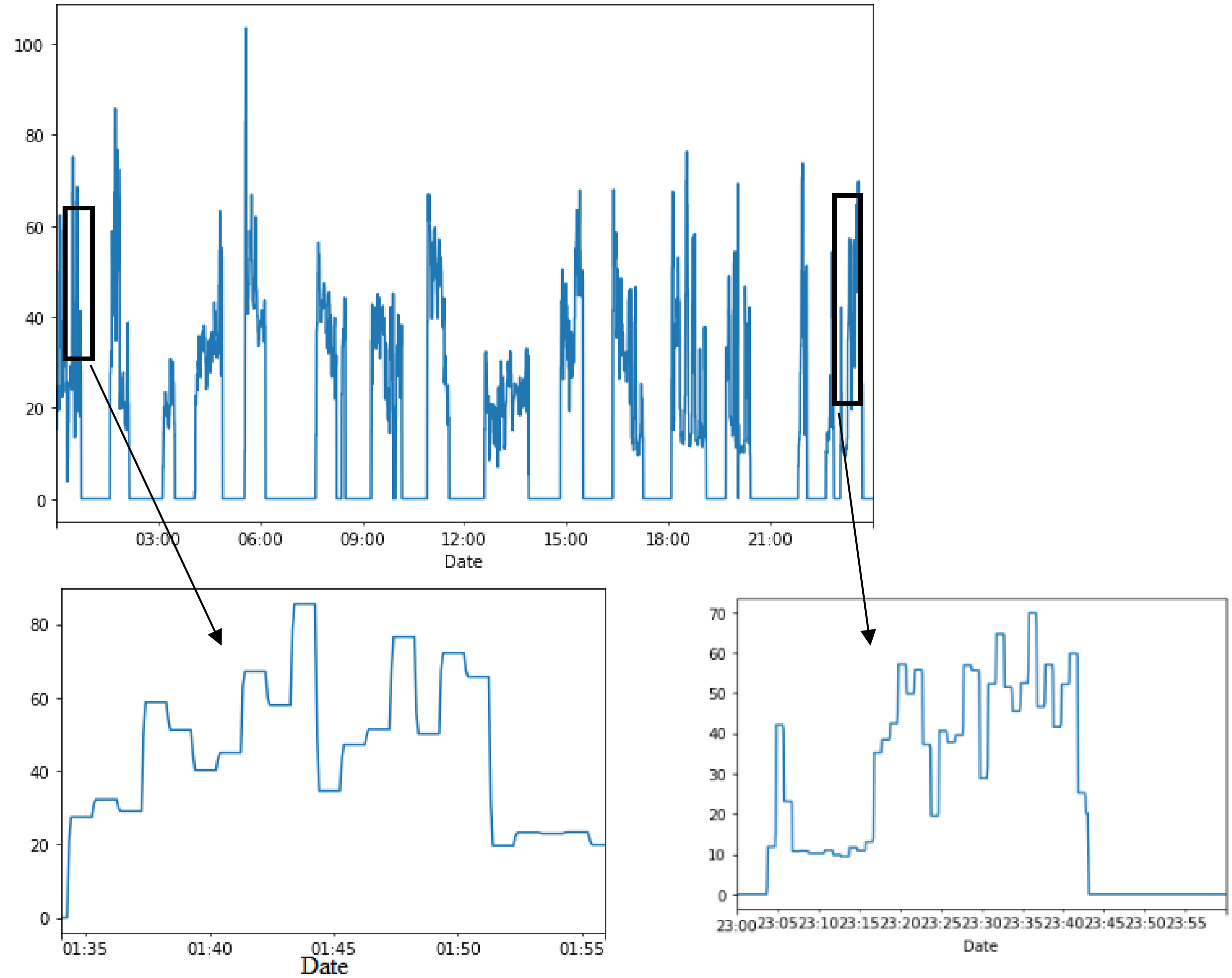


Figure 29:ROP variation with time

The figure 29 shows changes in the rate of penetration over time.

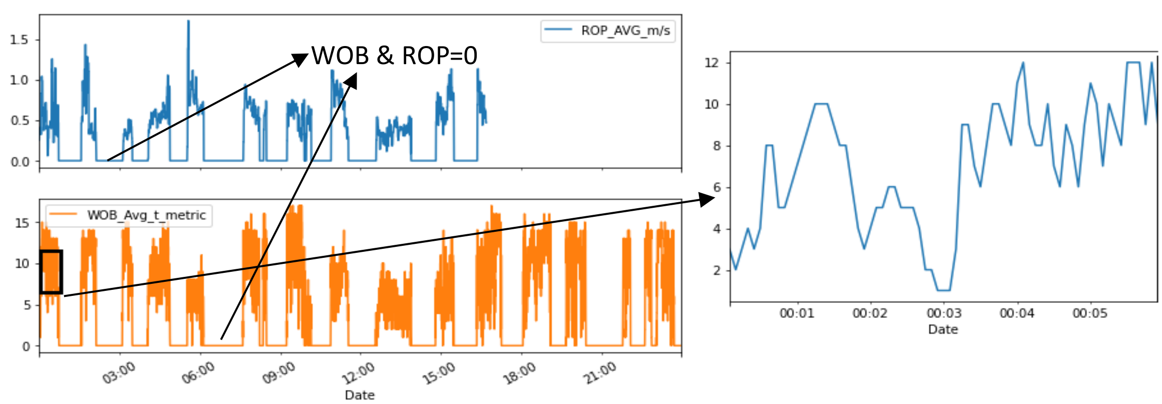


Figure 30:ROP vs WOB variation with time

when the value of WOB and ROP turns to zero, that's means that drilling stopped usually due to theaddition of drillpipe or circulation or to the change of bit (figure V-30)

V.2.3 correlation between the parameter:

Table 9: parameter correlation

	Depth_TVD_m	Bit_Position_m	ROP_AVG_h/m	ROP_AVG_m/s	WOB_Avg_t_metric	Torque_Avg_lbf.ft	RPM_Avg_rpm	Pressure_SPP_ Avg_psi	Flow_In_Pumps_(l/m)	Mud_Temperature_OUT_Avg_(°C)
Depth_TVD_m	1.000000	0.996906	-0.036448	0.011766	0.018074	0.038828	0.039382	0.146697	0.046261	0.826772
Bit_Position_m	0.996906	1.000000	0.004085	0.074614	0.060631	0.080354	0.074581	0.148712	0.043267	0.821006
ROP_AVG_h/m	-0.036448	0.004085	1.000000	1.000000	0.700917	0.773095	0.772850	0.433366	0.400700	-0.029427
ROP_AVG_m/s	0.011766	0.074614	1.000000	1.000000	0.775703	0.800313	0.805447	0.424920	0.394617	-0.005699
WOB_Avg_t_metric	0.018074	0.060631	0.700917	0.775703	1.000000	0.827892	0.815686	0.462636	0.419935	0.008765
Torque_Avg_lbf.ft	0.038828	0.080354	0.773095	0.800313	0.827892	1.000000	0.819892	0.487960	0.438237	0.098313
RPM_Avg_rpm	0.039382	0.074581	0.772850	0.805447	0.815686	0.819892	1.000000	0.707960	0.687903	0.100305
Pressure_SPP_Avg_psi	0.146697	0.148712	0.433366	0.424920	0.462636	0.487960	0.707960	1.000000	0.969168	0.203839
Flow_In_Pumps_(l/m)	0.046261	0.043267	0.400700	0.394617	0.419935	0.438237	0.687903	0.969168	1.000000	0.124751
Mud_Temperature_OUT_Avg_(°C)	0.826772	0.821006	-0.029427	-0.005699	0.008765	0.098313	0.100305	0.203839	0.124751	1.000000

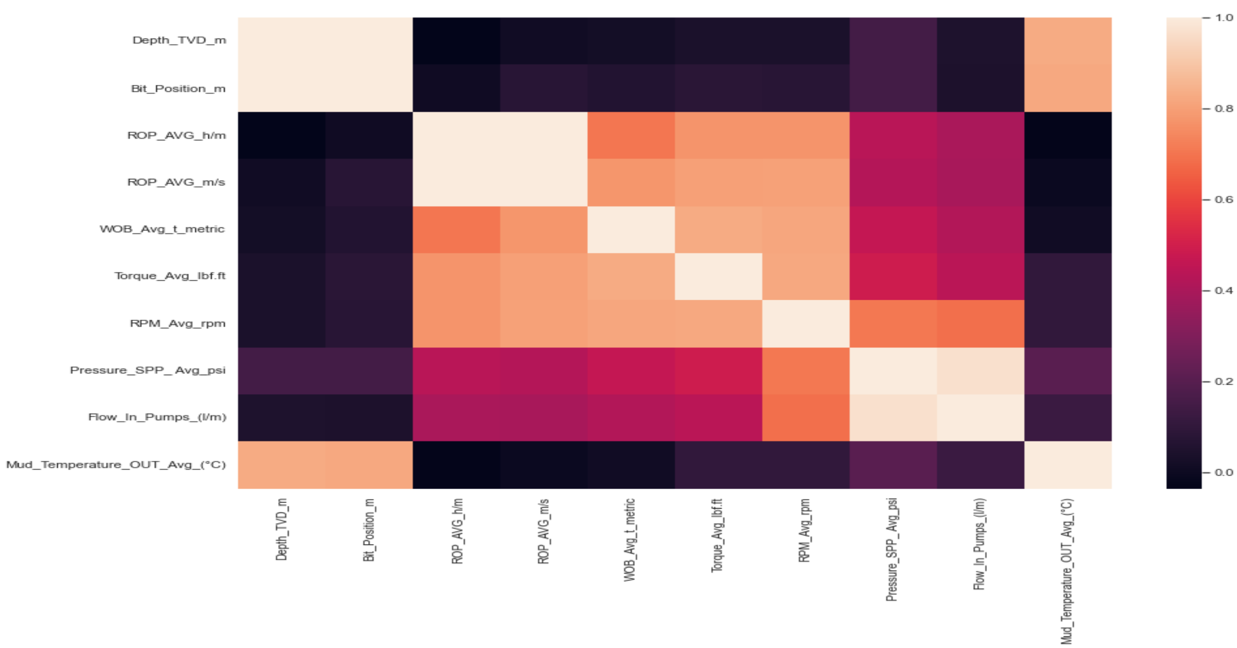


Figure 31:Parameter correlation

The figure 31 represent Correlation matrix for the performance parameters measured during drilling . The matrix below shows the correlation between each pair of parameters. A coefficient closer to (1) represents positive correlation and closer to (-1) represents negative correlation between the drilling variables.

we have significant correlation between rotation per minute (rpm) and torque with rate of penetration (0.80) .A high correlation was observed between temperature and depth, which is highly dependent on the temperature increase during drilling, while a low negative correlation was observed between ROP and depth(Table 9).

medelling result interpretation

This chapter will analyze the results of the modelling presented in chapter IV and review the reability of each modelling technique. This will be done through multiple methods to test how accurate the prediceted ROP**.**

V.3 Simple linear regression Algorithm:

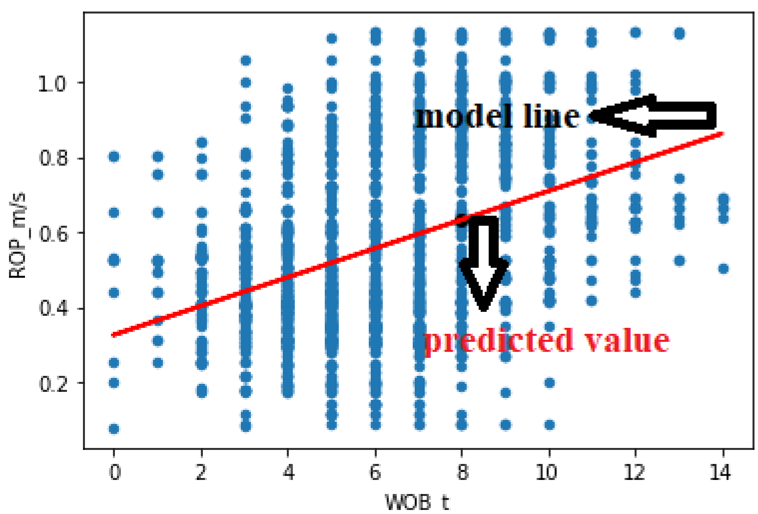


Figure 32: Scatter plot with regression model ROP vs WOB

$$\boxed{y = 0.325 + 0.038x} \quad \text{……………………….V.1}$$

```
data.plot(kind="scatter", x='WOB_t', y='ROP_m/s').
plt.plot(x, model.predict(x),color='red', linewidth=2)
plt.scatter(WOB_new, ROP_predict, color='black')
plt.show()
```

The y-intercept of 0.325 can be interpreted this way: when we don't have WOB, the ROP will be 0.325(m/s). The slope tells us that if it 8 (t) that the ROP would increase by an additional 0.629 m/s.(black point).

The regression line does not go through every point, instead it balances the difference between all data points and the straight-line model.it shows linear relationship

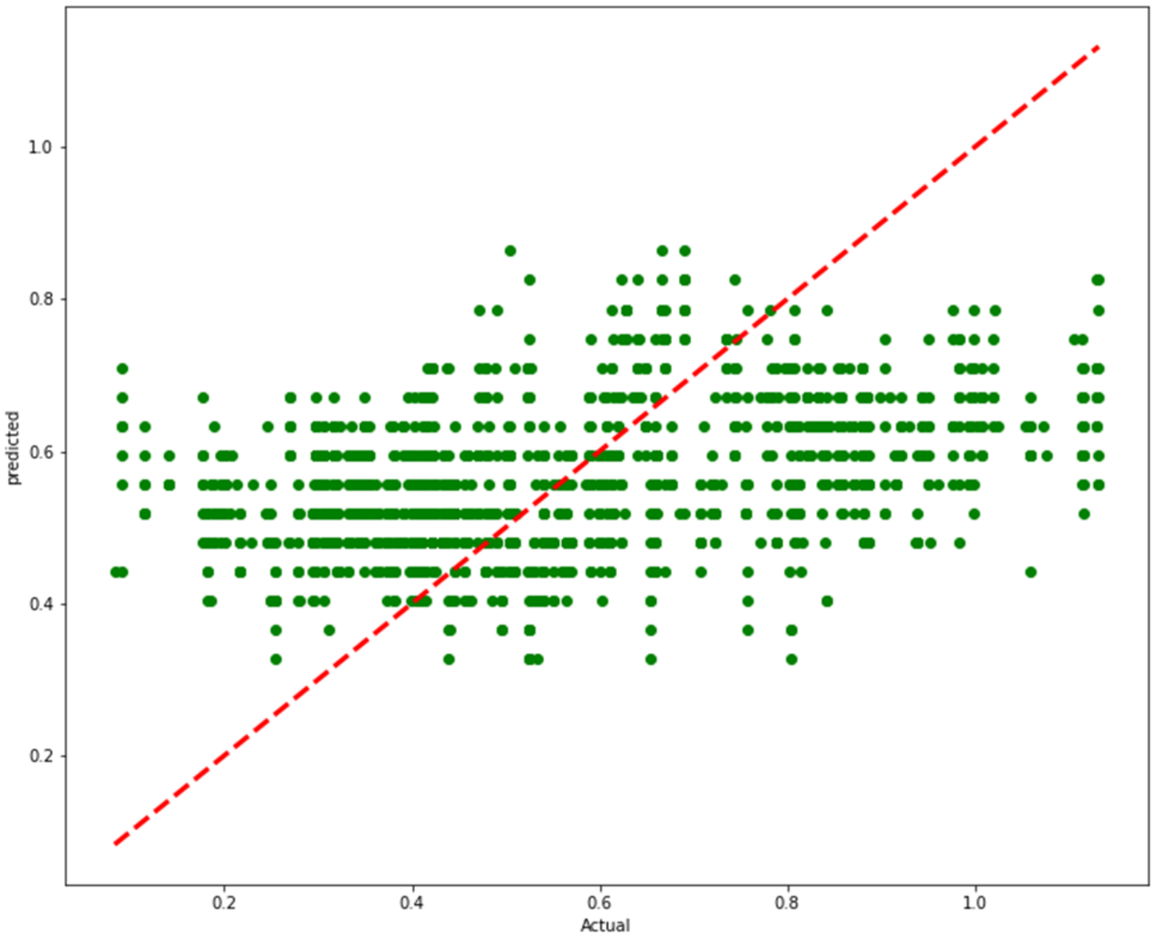


Figure 33: The model performance with WOB

The figure 33, shows the predicted values vs actuel values which the model line is linear and the point close to them that means the performance of the model is best. The red line represent the best fit line

The table V-8 represent the diffrence between predicte and actual values, when the difference is the smallest the prediction is the better.

Table 10: the difference between the actual and predicted values

weight on bit	observed ROP	predicted ROP	Diffrence
5	0.390000	0.518814	-0.128814
6	0.501667	0.557091	-0.055424
2	0.495000	0.403983	0.091017
10	0.478333	0.710199	-0.231865
5	0.858333	0.518814	0.339520
11	1.018333	0.748476	0.269858
2	0.495000	0.403983	0.091017
4	0.381667	0.480537	-0.098870
7	0.405000	0.595368	-0.190368
6	0.801667	0.557091	0.244576
4	0.881667	0.480537	0.401130
8	0.416667	0.633645	-0.216978
8	1.071667	0.633645	0.438022
3	0.381667	0.442260	-0.060593
6	0.676667	0.557091	0.119576
8	0.825000	0.633645	0.191355
10	0.835000	0.710199	0.124801
3	0.525000	0.442260	0.082740
9	0.778333	0.671922	0.106412
11	0.735000	0.748476	-0.013476
2	0.445000	0.403983	0.041017
13	0.525000	0.825030	-0.300030
8	0.921667	0.633645	0.288022
5	0.320000	0.518814	-0.198814
6	0.871667	0.557091	0.314576
4	0.788333	0.480537	0.307797
11	0.670000	0.748476	-0.078476
5	0.343333	0.518814	-0.175480
2	0.456667	0.403983	0.052684

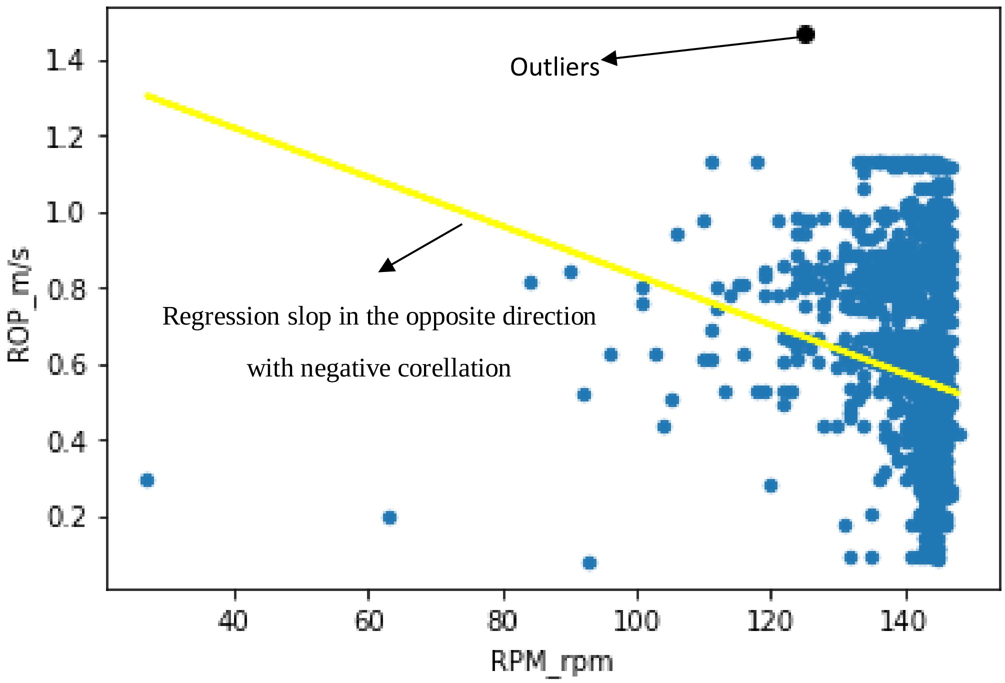


Figure 34:Scatterplot of regression model ROP vs WOB

$$\boxed{y = 1.477+ -0.00646x.} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{V.2}$$

b_1= -0.00646 is the model coefficient,that means very weak negative correlation.

The model shows that that the relationship between this two parameter it is not linear because of the opposite direction of the slop,the black point represent an outliers data.

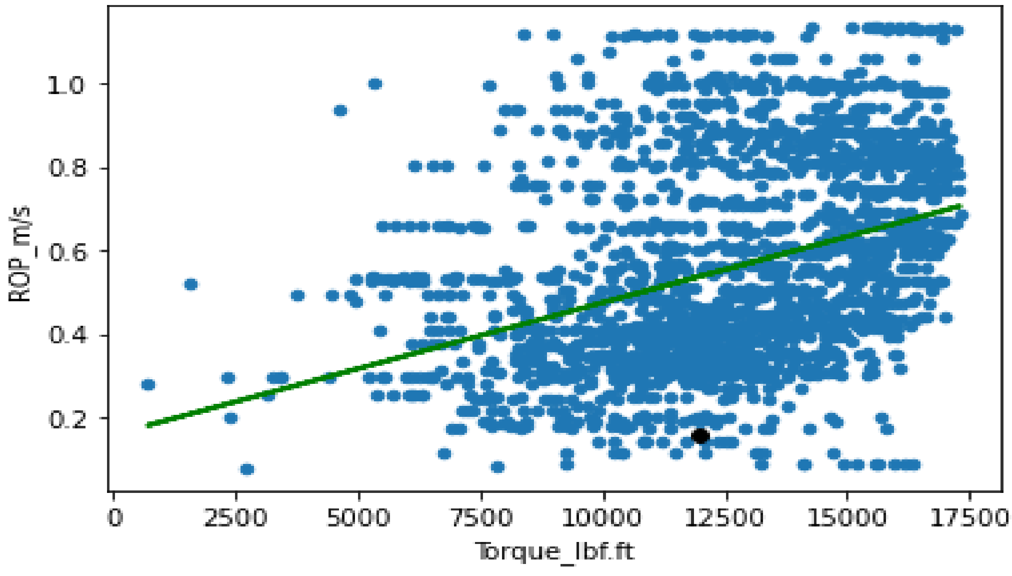


Figure 35: Scatter plot with regression model ROP vs torque

Model coefficient b1= 0.1587.

A scatter plot indicates that there is a weak positive relationship between ROP and torque (linear relationship).

Table 11: The model performance verification using a test value of torque.

	Torque_new	ROP_predicted_m/s
	0	0
0	13098	0.572509
1	14500	0.616800
2	14570	0.619011
3	15000	0.632595
4	15200	0.638914

Table 12: Comparison between actual and predicted values using multiple linear regression.

	actual values	predicted values	Diffrence
0	0.570000	0.517788	0.052212
1	0.943333	0.655456	0.287878
2	0.438333	0.632096	-0.193762
3	0.411667	0.610099	-0.198432
4	0.478333	0.666819	-0.188486
5	0.858333	0.564519	0.293814
6	0.490000	0.476675	0.013325
7	0.510000	0.514179	-0.004179
8	0.756667	0.436632	0.320035
9	0.565000	0.543256	0.021744
10	0.660000	0.396850	0.263150
11	0.983333	0.583647	0.399686
12	0.396667	0.488948	-0.092281
13	0.410000	0.535639	-0.125639
14	0.505000	0.634215	-0.129215
15	0.570000	0.634823	-0.064823
16	0.395000	0.591803	-0.196803
17	0.216667	0.467745	-0.251079
18	0.880000	0.626855	0.253145
19	0.690000	0.745359	-0.055359

The lowest the difference is the best accuracy will be.

V.4 The prediction result of multiple linear regression Algorithm:

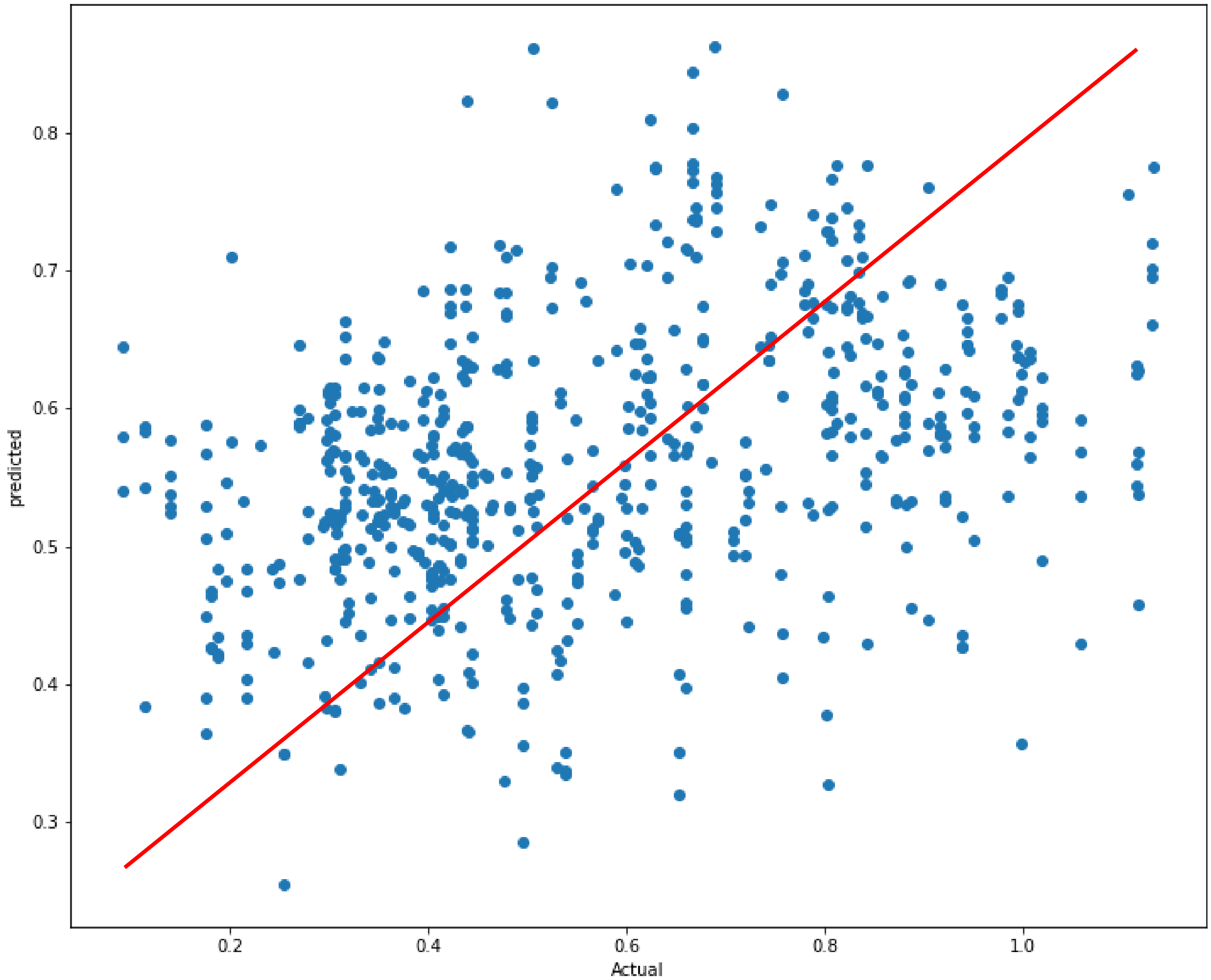


Figure 36: Scatter plot actual and predicted values using multiple linear regression

$$Y= 0.4121+2.232e\text{-}02\ x_1+1.743e\text{-}05\ x_2+(-1.477e\text{-}03)\ x_3 \quad \text{……V.3}$$

Y is the dependent variable (ROP) and x1, x2 and x3 are the independent variables (WOB, RPM, and torque). The values (2.232e-02, 1.743e-05 and -1.477e-03) are the coefficients of the model, the value 0.4121, is the intercept.

From the figure 36, we can observe that the model predictions are not so accurate and that maybe due to the quality of data or the chosen model.

Conclusion

Conclusion general

A drilling dataset from an oil well in HASSI MESSEOUAD was analyzed. As formation properties has a major impact on drill ability, The selection of the appropriate section was based on the lithological characteristics and data availability and cleaning and pre-processing of data was necessary before input it in the model. After completing this study, a real-time drilling data and model was achieved after optimization.

The aims of this work is to create a machine learning model for drilling optimization by maximizing ROP base on real-time drilling data and using simple and multilinear regression python models, and optimizing it by modifying controllable parameters.

A machine learning approach was successfully implemented to select data in the same range in well section during model development. After determining all the drilling parameters required to be used in rate of penetration model, Multiple Regression analysis is applied to obtain model constant.

The data such WOB, rotary speed (RPM) and torque are the independent parameters were used as input and ROP as dependent variable.

The Modeling and optimization were successfully completed using the machine learning approach, updating the model, and predicting and optimizing target parameter wish is the rat of penetration (ROP). The Machine Learning models were implemented (simple and multiple linear regression), and the simple linear regression algorithm was the best based on the scatter plot and code running time. During optimization, and the controllable drilling parameters constraints were chosen only to fit in the data range.

Finally, Results of simple and multilinear regression have shown a nonlinear reliance between the features and target of the ROP models. Many suggestions will be outlined below as future work:

1. Previous studies revealed that linear regression achieves remarkable results when using in ROP prediction. The limitation in this model was the time with more time on my hands, this relationship could definitely be interesting to look at. Thus, a more accurate model will reach better predictions and consequently more realistic optimization.

2. Data quality is of vital importance since this study is based on the relation between data. With more WOB and Torque corrections the model accuracy would definitively improve.

3. No specific constraints were selected during optimization, which would make a more realistic and complete model. These bounds can be modified based on drilling vibrations, hole cleaning, borehole stability, and pressure variation among other drilling issues.

Bibliographic

Bibliographic

[1] BESSAAD Farid GHRASLIA Haroun MEDIOUNI Mohamed Amine “Drilling parameters optimization : Application of modern optimization method” Faculty Of Hydrocarbons, Renewable Energy, Science Of The Earth and Universe, Kasdi Merbah University Ouargla, 15 / 06 / 2019

[2] Graham J.W. and Muench N.L., “Analytical Determination of Optimum Bit Weight and Rotary Speed Combinations,” SPE 1349-G, Fall Meeting of the Society of Petroleum Engineers, Dallas,October 1959.

[3] Galle E.M and Woods A.B., “Best Constant Weight and Rotary Speed for Rotary Rock Bits,” Drill. And Prod. Prac., pp 48-73, API 1963.

[4]Bourgoyne Jr., A. T. and Young Jr., F. S., “A multiple regression approach to optimal drilling and abnormal pressure detection,” Society of Petroleum Engineers Journal, August, 371-384.

[5] Schreuder J.C. and Sharpe P.J., “Drilling The Limit – A Key to Reduce Well Costs,” SPE 57258, Asia Pacific Improved Oil Recovery Conference, Malaysia, October 1999.

[6] “Introduction to Oil&Gas Well Drilling”(online). available://www.oil-gasportal.com/drilling/introduction-to-oilgas-well-drilling/

[7] Technical Area: 2 Drilling Tools and Equipment\ Teaching Guidance\Groundwater development and Management\Capacity Development Project\DDCAP

[8] “Components of a Land-Based Rotary Drilling Platform” online available: https://dtetechnology.wordpress.com/2014/05/04/components-of-a-land-based-rotary-drilling-platform/

[9] Drill String Components” online available: https://www.drillingcourse.com/2015/12/drill-string-components.html

[10] Components of the drill string\Oilfield Team \online available: https://oilfieldteam.com/en/a/learning/Components-of-the-drill-string

[11] Oil and Gas Well Drilling and Servicing eTool \online available: https://www.osha.gov/etools/oil-and-gas/illustrated-glossary

[12] Rotary drill rig components available: uomustansiriyah.edu.iq

[13] Kelly drive online available: https://en.wikipedia.org/wiki/Kelly_drive

[14] Drilling Mud Circulation System\online available: https://drillingfluid.org/drilling-fluids-handbook-2/drilling-mud-circulation-system.html

[15] Mud logging /Geothermal Power Generation/Developments and Innovation/Book /2016 by:Ronald DiPippo

[16] Mud Logging unit, mud Logging Cabin, Explosion Proof Pressurised Cabin, Atex Container, Explosion Proof Container/online available:https://www.tls-containers.com/tls-blog/mud-logging-unit-mud-logging-cabin-explosion-proof-pressurised-cabin-atex-container-explosion-proof-container

[17] Mud logging From Wikipedia,online available : https://en.wikipedia.org/wiki/Mud_logging

[18] Operational Manual for Mud Logging Engineers /oil Quest International ltd /Km 19 Port Harcourt Owerri Road,Igwuruta, port Harcourt/ Rivers State

[19] Internal report of training GRN

[20] Types of Logging /Well Logging /Mud logging/ 2012 All Star Training, Inc.

[21]Mud logging presentation university of AZAD JAMMU and KASHMIR/muzaff arabad institute of geology

[22] Real-Time-Optimization of Drilling Parameters During Drilling Operations/ The Degree of Doctor of Philosophy / Petroleum And Natural Gas Engineering/ February 2010

[23] TEDJINI ABDELREZZAK , HAMARHOUM SALAH EDDINE ,DJELLOUL FETHI/application of specific energy for bit selection HASSI MESSAOUD OMG-501/ university of KASDI MERBAH. OUARGLA

[24] Applications of Artificial Intelligence Techniques in Optimizing Drilling/ Mohammadreza Koopialipoor and Amin Noorbakhsh.

[25] Effect of Drilling Fluid Properties on Rate of Penetration/ Abouzar Mirzaei-Paiaman University of Campinas, M K Ghassem Al-Askari, B Salmani, B D Al-Anazi/ January 2009

[26] Mehaysen Mahasneh, Al-Balqa' Optimization Drilling Parameters Performance During Drilling in Gas Wells/ Article in International Journal of Oil Gas and Coal Engineering / Applied University/ March 2017

[27] Ozbayoglu M. E, Miska S. Z, Reed T, and Takach N,"Analysis of the Effects of Major Drilling Parameters on Cuttings Transport Efficiency for High-Angle Wells in Coiled Tubing Drilling Operations, " SPE 89334, SPE/IcoTA CT Conf. andExhb., Houston, TX, March 2004

[28] Al-Betairi E. A, Moussa M, and Al-Otaibi S, "Multiple Regression Approach to Optimize Drilling Operations in the Arabian Gulf Area, " SPE 13694, Middle East Oil Symposium, Bahrain, March 1985

[29] Application of Artificial Intelligence Techniques in Drilling System Design and Operations: A State of the Art Review and Future Research Pathways/ Opeyemi Bello, Institute of Petroleum Engineering, Clausthal University of Technology; Catalin Teodoriu, Mewbourne School of Petroleum and Geological Engineering, University of Oklahoma; Tanveer Yaqoob, Joachim Oppelt, Javier Holzmann, and Alisigwe Obiwanne, Institute of Petroleum Engineering, Clausthal University of Technology/Society of Petroleum Engineers 2016

[30] Understand Machine Learning and Its End-to-End Process/online: https://www.analyticsvidhya.com/blog/2020/12/understand-machine-learning-and-its-end-to-end-process

[31] A review on significance of sub fields in AI/ International journal of latest trends in engineering and technology(IJLTET)/ Dr. Viji Priya, Dr.Jammi Ashock, Dr.S Supiach

[32] Application of machine learning and artificial intelligence in oil and gas industry, Petroleum Research, Volume 6, Issue 4, December 2021, Pages 379-391

[33] Natural Language Processing Applications/ online available: https://julenetxaniz.eus/en/project/nlp-applications/

[34] Fuzzy Logic Applications in Chemical Processes/online available: https://www.semanticscholar.org/paper/Fuzzy-Logic-Applications-in-Chemical-Processes-Emami/

[35] An Overview on Applications of Machine learning in petroleum Engineering/3rd International Congress on Science and Engineering/ Mohammad Hossein Motamedie , Farshad jafarizadeh/ HAMBURG – GERMANY March 2020

[36]Machine Learning Simplified/online available: https://randerson112358.medium.com/machine-learning-simplified-

[37] What is CRISP DM?/online available: https://www.datascience-pm.com/crisp-dm-2/

[38] Data mining Assignment Paper/online available : https://xpertessay.com/2022/05/20/data-mining-assignment-paper-2/

[39] Review on the Application of Machine Learning Algorithms in the Sequence Data Mining of DNA/Front. Bioeng. Biotechnol., 04 September 2020

[40] Data Preprocessing in Data Mining/online available: https://www.geeksforgeeks.org/data-preprocessing-in-data-mining/

[41] Online available : https://jupyter.org/

[42]Online available : https://pythonnumericalmethods.berkeley.edu/notebooks/chapter01.04-Introduction-to-Jupyter-Notebook.html

[43] Online available: https://www.mygreatlearning.com/blog/open-source-python-libraries/)

[44] Breiman, L., 2001b. Statistical modeling: the two cultures (with comments and a rejoinder by the author). Stat. Sci. 16, 199–231.

[45] Malik Alsenwar,' NCS Drilling Data Based ROP Modelling and its Application', Petroleum Engineering/Drilling Engineering, FACULTY OF SCIENCE AND TECHNOLOGY, 2017

[46] Chapter 7: Correlation and Simple Linear Regression/online availaible: https://milnepublishing.geneseo.edu/natural-resources-biometrics/chapter/chapter-7-correlation-and-simple-linear-regression/

[47] The Linear Regression Model,online avalaible: https://www.econ-analysis.com/single-post/2016/05/27/the-linear-regression-model-1

عنوان المذكرة: استكشاف بيانات الحفر في الوقت الفعلي لتحسينها باستخدام التعلم الآلي

الاسم: فرح فاطمة الزهراء   اللقب: عزاوي   إشراف: أسماء حجاج

الملخص

يعد الحصول على الحد الأقصى لمعدل الاختراق (ROP) أحد الأساليب العديدة لتقليل التكلفة والوقت غير الإنتاجي (NPT) في آبار حفر النفط والغاز. تؤثر العديد من المعلمات على معدل الاختراق، بما في ذلك تنظيف الفتحات، وتآكل الأسنان، وما إلى ذلك. تم تطوير الدراسة في أربعة أجزاء. أولاً، لدينا عمومية الحفر وتسجيل المعلمات. في الجزء الثاني، نماذج الذكاء الاصطناعي والتعلم الآلي (ML). ثالثًا، كيفية المعالجة المسبقة للبيانات وتنظيفها وتحضيرها قبل الاستخدام. أخيرًا، تم تنفيذ عرض البيانات وتفسيرها ونماذج التعلم الآلي (الانحدار الخطي البسيط والمتعدد) لإنشاء نموذج ROP وتم اختيار القسم الذي يحتوي على أفضل أداء.
في هذه الدراسة، تم نمذجة وتحليل معلمة واحدة (ROP). يستخدم النموذج عمق البت والوزن على البت (WOB) وسرعة الدوران (RPM) وعزم الدوران (T) كمدخلات لإجراء الانحدار والتنبؤ بـ ROP. الهدف من هذا العمل هو دراسة إمكانية تطبيق خوارزميات التعلم الآلي المختلفة للتنبؤ بتحسين بعض معلمات حفر الآبار لتجنب المشاكل وتوفير الوقت والمال

Memory Title: Exploration of real-time drilling data for optimization using machine learning

Name: AZZAOUI   First Name: Farah Fatmazohra   Directed: Asma HADJADJ

Abstract

Obtaining the maximum rate of penetration (ROP) is one of many techniques to reduce cost and Non-Productive Time (NPT) in oil and gas drilling wells. Many parameters affect ROP, including hole cleaning, tooth wear, etc. The study was developed in four parts. First, we have generality of drilling and parameter recording. In the second part, artificial intelligence and machine learning (ML) models. Third, how to pre-process data, clean and prepare before use. Finally, data presentation and their interpretation, machine learning (ML) models (simple and multiple linear regression) were implemented to create a ROP model and the section with the best performance was selected. In this study, one parameter (ROP) was modelled and analyzed. The model uses Bit depth, Weight on Bit (WOB), rotary speed (RPM) and torque (T) as inputs to make a regression and predict ROP. The objective of this work is to study the applicability of various Machine Learning algorithms to predict optimizing some well drilling parameters to avoid problems and save time and money

Titre du mémoire :

Nom : AZZAOUI   Prénom :Farah Fatmazohra   Encadreur : Asma HADJADJ

Résumé

L'obtention du taux de pénétration maximal (ROP) est l'une des nombreuses techniques permettant de réduire les coûts et le temps non productif (NPT) dans les puits de forage pétrolier et gazier. De nombreux paramètres affectent la ROP, notamment le nettoyage des trous, l'usure des dents, etc. L'étude a été développée en quatre parties. Tout d'abord, nous avons la généralité du forage et l'enregistrement des paramètres. Dans la deuxième partie, les modèles d'intelligence artificielle et d'apprentissage automatique (ML). Troisièmement, comment pré-traiter les données, les nettoyer et les préparer avant utilisation. Enfin, la présentation des données et leur interprétation, des modèles d'apprentissage automatique (ML) (régression linéaire simple et multiple) ont été mis en œuvre pour créer un modèle ROP et la section la plus performante a été sélectionnée. Dans cette étude, un paramètre (ROP) a été modélisé et analysé. Le modèle utilise la profondeur de bits, le poids sur les bits (WOB), la vitesse de rotation (RPM) et le torque (T) comme entrées pour effectuer une régression et prédire le ROP. L'objectif de ce travail est d'étudier

l'applicabilité de divers algorithmes d'apprentissage automatique pour prédire l'optimisation de certains paramètres de forage de puits afin d'éviter les problèmes et d'économiser du temps et de l'argent