ALGERIAN DEMOCRATIC PEOPLE'S REPUBLIC

Ministry of Higher Education and Scientific Research

N° Serie: ………. /2024

University Kasdi Merbah Ouargla

Faculty of hydrocarbons, renewable energies and science of the earth and the universe

DEPARTMENT OF DRILLING AND MECHANICAL OIL FIELD

MEMORY

To Obtain the Master's Degree

Option: Oil drilling

Presented By:

CHERADID Abdel Bari, BEN SALEM Zahra, FRIHI Abdennour

-THEME-

Computational prediction of the drilling rate of penetration (ROP): A comparison of various machine learning approaches and traditional models

Defended on: / / 2024 before the examination committee

Jury:

President:	HADJ SAAD NADIR
Examiner :	CHAOUICHA SAMIRA
Supervisor:	KHELIFA ECHERIF

وما توفيقي إلا بالله

Thanks

All gratitude and thanks to God our creator who gave us the strength to carry out and complete this work.

We would like to firstly and warmly thank

Our promoter Mr. Khelifa Cherif

For agreeing to lead our work, for his precious advice, for his openness and availability.

Thanks to him, our work was done.

We also thank: Mr. Merabti Hocine and Mr. Chetti Djamel Eddine

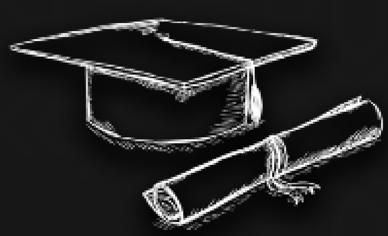
For the trust they have always shown in us and for their constant encouragement.

We thank everyone who helped us directly or indirectly. to finalization of this work,

We would like to express our sincere gratitude to them.

Finally, our thanks to all the teachers who contributed to our training at the

University of Ouargla.



Dedication

With great pride and honor, and after a long journey filled with hard work, diligence, and perseverance,

I stand today on the threshold of a new phase in my life to celebrate my graduation from university.

This achievement could not have been realized without the tremendous efforts and significant sacrifices made throughout the years of study,

the long hours of staying up late, studying, and researching. In this joyous moment, I cannot but extend my deepest thanks and gratitude to the dearest people to my heart, my beloved parents.

To you, who have been my support and backbone throughout every stage of my life.

To all my teachers, and colleagues, thank you for every moment we spent together, for the support and encouragement, for the advice

and beautiful memories that will remain etched in my memory forever.

This graduation is not the end of the road but a new beginning for another journey of striving

for excellence and achieving dreams.

Abdelbari

Dedication

I thank God very much for enabling me to complete this work

after years of education and effort

I want to dedicate my work, first and foremost, to my dear mother, who gave me life in its entirety.

Thank you, my lady, and to my beloved father

Your wisdom, guidance, and belief in my abilities have pushed me forward.

I thank you because you are the pillar of my strength

and my constant source of inspiration.

To my dear friends

I thank everyone, Shaima, Sajida, and Marwa, for your friendship, constant support, and standing with me through thick and thin.

To my beloved ones

You are my first friends, those close to me, and my successors.

I thank my big sisters, Sarah and Malika, for being my constant companions on this journey called life.

Zahra

Dedication

thank God very much for enabling me to complete this work

I want to dedicate my humble work to all my family members, especially to my dear mother,

who supported me throughout my academic journey. I thank you for your tremendous efforts.

I thank my dear father

for his sacrifices and constant perseverance.

I thank him for the virtuous values and morals that he instilled in me.

I thank my brothers for

being my constant companions

and supporters on this journey, and

I also thank my friends for

sharing all my joys

and sorrows, and you are the family that

I chose after my family.

Abdennour

Table of contents

CHAPTER II: Drill Bit Selection

CHAPTER III: Artificial intelligence and its applications in oil field

List of tables

List of figures

Nomenclature

ROP : Rate of penetration

PDC : Polycrystalline diamond compact

TSP : Thermally Stable Polycrystalline

IADC : International Association of Drilling Contractors

AI: Artificial Intelligence

TCI : tungsten carbide insert

BHA : Bottom hole assembly

ML : Machine learning

KNN : K-Nearest Neighbors

XGBoos : Extreme Gradient Boosting

ANN : artificial Neural Network

GLM : Generalized linear model

SVR : support vector regression

GPR : Gaussian process regression

API : Application programming interface

NPT : Non-productive time

TAGS : Triassic Argileux Greseux Superieur

TAGI : Triassic Argileux Greseux Inferior

TD : total depth

TVD : total vertical depth

WOB : weight on the bit

RPM : Rtation per minute

SPP : Standpipe pressure

RMSE : Root Mean Squared Error

$: Dollar

Hr : Hour

$/hr : Doller per hour

Ft/hr : Foot per hour

ملخص : هذه المذكرة تستكشف عدة محاور مهمة في صناعة النفط والغاز، حيث تركز على تقنيات الحفر واختيار الأدوات المناسبة، بالإضافة إلى استخدام الذكاء الاصطناعي في تحسين العمليات الاستكشافية والإنتاجية في حقول النفط.كما تناولت المذكرة معلومات عامة عن أدوات الحفر وأبرزت الدور الحاسم الذي تلعبه في عمليات الحفر الناجحة، كما استعرضت تعقيدات اختيار الأدوات المناسبة لظروف الحفر المختلفة.كما قدمت المذكرة استكشافاً لتطبيقات الذكاء الاصطناعي في صناعة النفط والغاز، مشيرة إلى الفرص الكبيرة التي تتيحها هذه التقنيات لتحسين العمليات الإنتاجية وتقليل التكاليف، من خلال تحليل البيانات الضخمة وتنفيذ خوارزميات التعلم الآلي.وفي النهاية، تناولت المذكرة الإنتاجية المتعلقة بمعدل الاختراق واختيار أدوات الحفر بناءً على هذا المعدل، موضحة أهمية فهم ديناميات الحفر واستخدام بشكل عام، تبرز هذه المذكرة أهمية التقنيات المتقدمة مثل الذكاء الاصطناعي .مؤشرات الأداء لاتخاذ القرارات الصائبة في تحسين كفاءة صناعة النفط والغاز وتعزيز استخراج الموارد بطرق مستدامة.

الكلمات المفتاحية : تنقيب – اداة حفر- ذكاء اصطناعي- تعليم الالة – تنبؤ – خوارزمية – نموذج .

Abstract: This thesis explores several important aspects of the oil and gas industry, focusing on drilling techniques and selecting appropriate bits, as well as the use of artificial intelligence to enhance exploration and production operations in oil fields. The thesis also provides general information about drilling bits and highlights the crucial role they play in successful drilling operations, as well as the complexities of selecting the right tools for various drilling conditions. Additionally, the thesis delves into the exploration of artificial intelligence applications in the oil and gas industry, indicating the significant opportunities these technologies offer to improve production processes and reduce costs through big data analysis and the implementation of machine learning algorithms. Furthermore, the thesis addresses productivity related to the rate of penetration and the selection of drilling tools based on this rate, emphasizing the importance of understanding drilling dynamics and using performance indicators to make informed decisions. Overall, this thesis underscores the importance of advanced technologies such as artificial intelligence in enhancing the efficiency of the oil and gas industry and promoting sustainable resource extraction methods.

Key word : Drilling- Drilling bit – Artificial intelligence – Machine learning – Prediction- Algorithm – Model.

Résumé : Ce mémoire explore plusieurs aspects importants de l'industrie pétrolière et gazière, en mettant l'accent sur les techniques de forage et la sélection d'outils appropriés, ainsi que sur l'utilisation de l'intelligence artificielle pour améliorer les opérations d'exploration et de production. De plus, le mémoire donner des informations générales sur les outils de forage et leurs rôle qu'ils jouent dans les opérations de forage réussies, ainsi que les complexités du choix des bons outils pour différentes conditions de forage. De plus, le mémoire explore les applications de l'intelligence artificielle dans l'industrie pétrolière et gazière, indiquant les importantes opportunités offertes par ces technologies pour améliorer les processus de production et réduire les coûts grâce à l'analyse des données et à la mise en œuvre d'algorithmes d'apprentissage automatique. Enfin, le mémoire aborde la productivité liée au taux de pénétration et à la sélection des outils de forage en fonction de ce taux, en soulignant l'importance de comprendre la dynamique du forage et d'utiliser des indicateurs de performance pour prendre des décisions. Dans l'ensemble, ce mémoire met en évidence l'importance des technologies avancées telles que l'intelligence artificielle pour l'amélioration de l'efficacité de l'industrie pétrolière et gazière et la promotion de méthodes durables d'extraction des ressources.

Mots clé : Forage – l'outil de forage – Intelligence artificielle – apprentissage automatique – Prédiction – Algorithme – modèle .

General Introduction

This thesis aims to explore and discuss several pivotal aspects related to the oil and gas industry, focusing on modern drilling and extraction techniques. It encompasses three main axes: general information on drill bits, drill bit selection, and the application of artificial intelligence in oil fields.

Drill bits are among the most crucial elements in the process of oil and gas extraction. Available in a variety of designs and materials, selecting the appropriate tool is vital to ensure the success of the drilling operation. This thesis will review in

Chapter 1: The different types of drill bits and their characteristics, along with the factors to consider when choosing the right bit for specific drilling conditions.

Chapter 2 : The traditional method to selecting drilling bit .

Chapter 3 : Given the tremendous advancements in technology, artificial intelligence represents an intriguing field of interest in the oil and gas industry. AI technologies offer significant opportunities to enhance exploration and production processes, reduce costs, and increase production efficiency. Additionally,

Chapter 4 : will delve into modern AI applications in oil fields, including the utilization of big data and its analysis, and the implementation of machine learning techniques to improve drilling operations and by Improving rate of penetration.

In summary, this thesis looks forward to exploring and discussing how advanced drill bit tools and AI techniques can enhance the efficiency of the oil and gas industry, improving exploration and production processes.

CHAPTER I:

General information on bits

1 Introduction:

Drilling bit is the smallest but most important part of building an oil well. It is the first that comes into contact with the formations traversed under the action of an axial load and driving torque.

The cheapest bit is not necessarily the one with which you can make progress the fastest. The usage and operating times must be taken into account.

The destruction of the rock and the speed of advancement essentially depend on the type of tool and the drilling parameters applied to it.

Drilling parameters can be divided into two categories:

- **Mechanical parameters**:

1. Bit type.
2. Weight on the bit.
3. Rotation speed.

- **Hydraulic parameters:**

1. Flow rate.
2. Type of drilling fluid and its properties (density, viscosity and filtration).

Drilling bits have evolved over time to respond to the increasingly complex technical problems involved in drilling well construction.

The aim of all of these developments was to increase the feed rate and tool life and thus reduce drilling costs.

Drilling bits can be divided into two groups:

1. Roller bits.

2. Diamond bits.[3]

2 Objectives:

By the end of this Section, you should be able to:

• Recognize different bit types.

• Describe various design considerations for roller cone and PDC Bits.

• Select bits for various formation types and drilling conditions.

• Grade bits using the IADC Dull Grading System

• Utilize the IADC code to describe and compare bits.

• Identify important operational aspects that effect bit performance [2]

3 Drill Bit Types:

- Roller Coin Bits
- Diamond Bits

3.1 Roller cone bits :

Roller cone bits are a common type of drilling bit used in oil and gas well drilling operations. These bits consist of a metal body with several rolling cone-shaped cutters equipped with cutting teeth that rotate when subjected to pressure and friction from rock formations during drilling. Roller cone bits come in a wide range of designs and sizes, allowing engineers to select the most suitable type for specific drilling conditions and the anticipated rock properties. The main types of roller cone bits include milled tooth bits and insert tooth bits. (Figure I.01) Milled tooth bits have steel teeth milled directly into the cones, while insert tooth bits have tungsten carbide inserts pressed into the cone surfaces for enhanced durability and performance. [2]

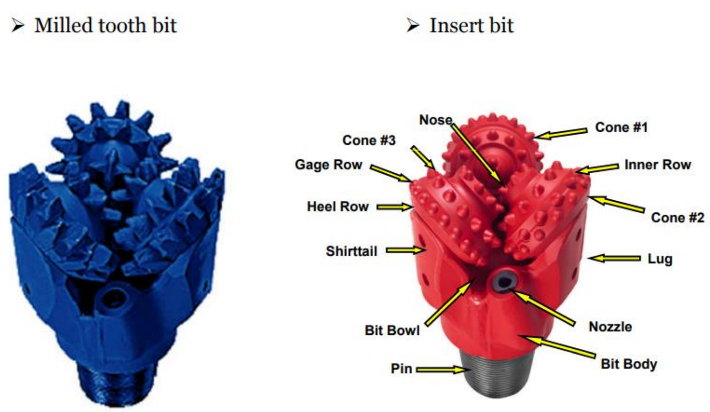


Figure I.01: Roller cone bits. [4]

3.2 Diamond Bits:

Diamond bits, also known as diamond core bits, are specialized drilling tools used primarily in hard and abrasive rock formations. These bits feature industrial-grade diamond segments embedded into the cutting face, providing exceptional hardness and abrasion resistance.

Diamond bits are classified into two main types: impregnated diamond bits and surface set diamond bits. Impregnated diamond bits have diamond particles embedded throughout the matrix of the bit, offering high durability and longevity. Surface set diamond bits have individual diamond crystals set on the bit surface through a metal matrix, providing aggressive cutting action in extremely hard formations.

Both types of diamond bits are widely used in geotechnical and mineral exploration drilling, as well as in the oil and gas industry for drilling through challenging formations.[2]

a) Natural Diamond Bits:

The hardness and wear resistance of diamond made it an obvious material to be used for a drilling bit. The diamond bit is really a type of drag bit since it has no moving cones and operates as a single unit. Industrial diamonds have been used for many years in drill bits and in core heads (Figure I.02).

The cutting action of a diamond bit is achieved by scraping away the rock. The diamonds are set in a specially designed pattern and bonded into a matrix material set on a steel body. Despite its high wear resistance diamond is sensitive to shock and vibration and therefore great care must be taken when running a diamond bit. Effective fluid circulation across the face of the bit is also very important to prevent overheating of the diamonds and matrix material and to prevent the face of the bit becoming smeared with the rock cuttings (bit balling).

Figure I.02: Natural Diamond bits.[4]

b) PDC Bits:

A new generation of diamond bits known as polycrystalline diamond compact (PDC) bits were introduced in the 1980`s (Figure I.03). These bits have the same advantages and disadvantages as natural diamond bits but use small discs of synthetic diamond to provide the scraping cutting surface. [6]

Figure I.03: PDC (Poly crystalline Diamond Compacts) bits. [3]

c) TSP Bits:

A further development of the PDC bit concept was the introduction in the later 1980's of Thermally Stable Polycrystalline (TSP) diamond bits. These bits are manufactured in a similar fashion to PDC bits but are tolerant of much higher temperatures than PDC bits. [7] (Figure I.04)

Figure I.04: TSP (Thermally Stable Poly crystalline) bits. [3]

4 Bit design:

4.1 Journal angle, cone profile:

One of the main design features of roller cone bits is journal angle. The journal angle is the angle formed by an axis of the journal relative to a horizontal plane. (Figure I.05)

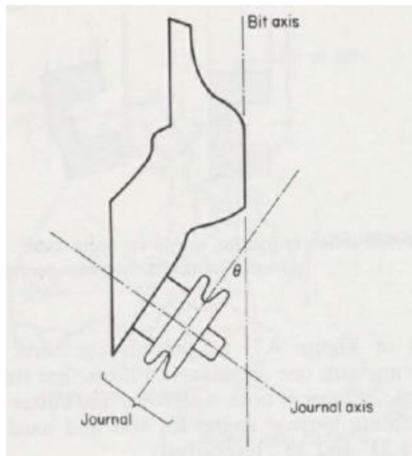


Figure I.05: Journal angle [2]

There is a close relationship between cone profile and stability of the bit. Cones with rounded profile provide a faster ROP, but are more labile. While cones with more flat profile are more durable, yet deliver lower penetration.

The journal angle has a direct influence on the size of the cone, with its growth the cone size declines.[2]

4.2 The journal angle depends on the type of rock:

- Soft formations – (journal angle 330) – allows greater penetration of the formation
- Medium formations – (journal angle 340 – 360) – decrease of cutter action
- Hard formations – (journal angle 390) – further decrease of cutter action

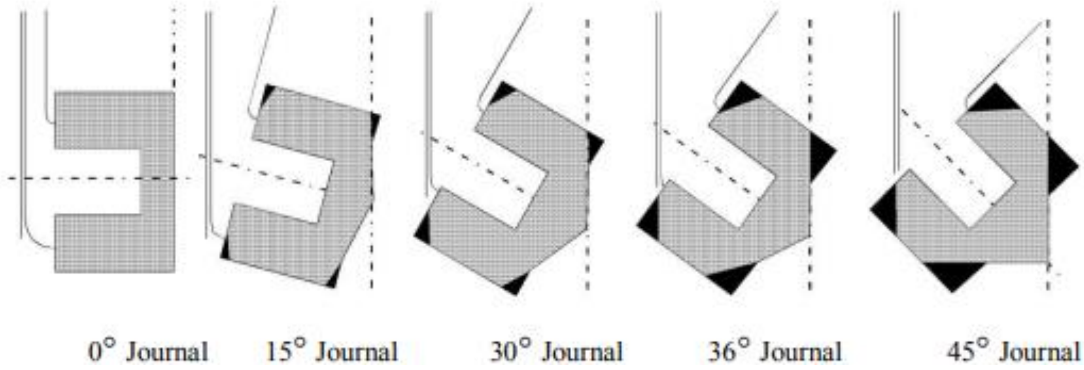


Figure I.06: Journal angles in roller cone bits [4]

4.3 Cone offset:

The "offset" specifies to a certain degree a drilling action of the roller cone bit (Figure I.07) illustrates cone offset. Shift of the cone's axis to the centerline of the bit is defined as "offset". The roller cone bit with no offset has the intersection point of cones axis in the center of the bit. The size of offset depends on the type of rock to be drilled. Its values range from 40 for soft formations to 0 for hard formations. Angular measure of the offset is called skew angle.

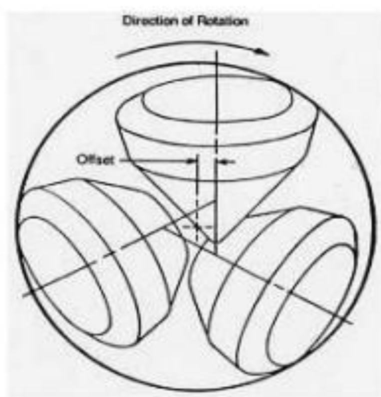
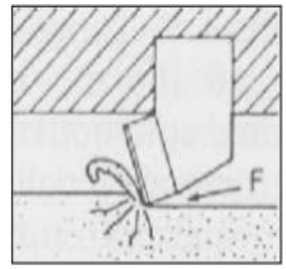


Figure I.07: Cone offset.[4]

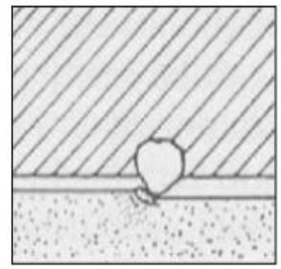
The cone offset results in interim stops in rotation and break the hole like a drag bit. With increasing the offset the bit wear increases proportionally.

4.4 Cutting Mechanisms:

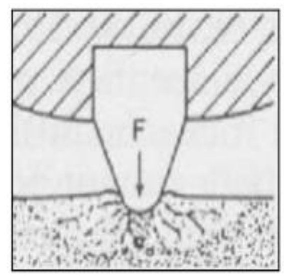
a) Shearing the formation : PDC bits use a shearing action to cut the rock. The polycrystalline diamond cutters, which are very sharp and hard, scrape and shear the rock as the bit rotates. TSP bits may have cutters shaped similarly to PDC bits, but they are optimized for thermal stability.



b) Ploughing / Grinding the formation : The weight on the bit (WOB) and the rotational speed (RPM) are critical factors. The correct combination ensures that the cutters maintain an optimal contact with the formation for effective shearing.



c) Crushing the formation by putting the rock in compression as a roller bit as the bit rotates, the cutters continuously contact the rock, maintaining a constant shearing action.



5 IADC (International Association of Drilling Contractors):

5.1 IADC Classification of Roller Cone Bits:

Roller cone bit classification system consists of four characters (three digits and a letter) and allows to classify all roller cone bits with milled teeth and inserts.

The first character: a digit from 1 to 8 (Table I.01)

- a digit from 1 to 3: roller cone bit with milled teeth (soft formations).

- a digit from 4 to 8: roller cone bit with inserts (hard formations).

Table I.01: Classification by type of formation of Roller Cone Bit [1]

The first number	Type of formation
1	Very soft formations
2	Medium soft formations

3	Soft formations
4	Very slightly hard formations
5	Slightly hard formations
6	Medium hard formations
7	Hard formations
9	Very hard formations

The second character: A digit from 1 to 4 that designates the hardness of each formation group It is a sub-classification of the formation hardness in each of the 8 classes determined by the 1st digit.

The third character: A digit from 1 to 9 that designates the characteristics of the roller cone bit.

The fourth character: a letter of the alphabet that defines additional characteristics of the roller cone bits.

The fifth character: sealed roller bearing with insert gauge protection.

5.2 IADC Classification of Diamond Bits:

A four-character coding system allows to classify all diamond bits.

The first character: Type of diamonds and matrix (Table I.02)

Table I.02: Classification by type of diamonds matrix of Diamond Bit.[1]

The first character	Type of diamonds matrix
D	natural diamond bit with a tungsten
M	PDC bit with a tungsten carbide matrix
S	PDC bit with a steel matrix

L	TSP bit with a tungsten carbide matrix;
O	For other types of fluid outlet

The second character: Bit profile: the bit profiles are coded using nine digits from 1 to 9 that represent the usual shapes of diamond bits.

The third character: Hydraulic characteristics it is a code using character that represent the hydraulic characteristics of diamond bits. (Table I.03)

Table I.03: Classification by hydraulic characteristics of Diamond Bit.[1]

The third character	Hydraulic characteristic's
R	for radial fluid outlet
X	for central groove outlet
O	for other types of fluid outlet

The fourth character: Size and density of diamonds: nine digits (1 to 9) symbolize the type, size and density of diamonds. The size of natural diamonds is based on the number of stones per carat, while the size of synthetic diamonds is based on the dimension of the cutting edge. The diamond density gives a relative indication of their number to distinguish the heavily loaded bits from the lightly loaded ones).

CHAPTER II:

Drill Bit Selection

1 Introduction:

The optimal drill bit series, optimizing operating parameters, and identifying bit-related drilling risks are all key factors in determining consistent well construction performance at the lowest possible cost per foot. With many choices available today, selecting the proper bit for a particular application can become quite confusing. There is a long list which includes, spade, drag and shear type bits, point attack bits, PDC bits, milled tooth and TCI (tungsten carbide insert) roller cone bits. There are IADC codes, soft, medium, and hard categories for each bit, sealed, non-sealed, roller, and friction bearings. The real job to be done is choosing the appropriate bit which will help to optimize the ROP of the operations thus enhancing the technical limit value. As stated, before the wrong method of bit selection is often used by operators where the operator packages data (log data, offset dull grading etc.) and gives this data to a drill bit contractor who uses this data to select the bit he thinks is most appropriate to drill the section. Most times the choices made could be very wrong and will impact on the cost of the well. This cost impact could be very high if the ROP reduces and a trip is needed out of hole to change the drill bit. [8]

2 Importance of study:

Selection is one of activities the drilling engineer needs to do. the drilling engineer which himself often isn't an expert on bit design has to relay on the recommendations from the bit manufacturer expert. However, to choose between the different bit manufacturers recommendations can be hard. Often selection is based on previously bit manufacturers experience in the field or based on the bit manufacturers success with a new design elsewhere in the world. The decision criteria for selection bit are often vague. And after the decision is made it is hard for the drilling engineer on an objective basis justifying the bit selection, both to the internal organization and also to the other bit companies who lost the tender. Oil companies have chosen different approaches to try to overcome this problem. In some cases, the service companies have got the responsibility for bit selection by engaging the service companies' optimization engineers.

3 Cost per foot:

The basic form of bit selection is normally done based on cost per foot. This method is simply choosing the bit that will provide the lowest cost per foot over the upcoming interval. In addition to that, other factors are taken into consideration as well such as offset, journal angle, and other design aspects. This differentiates one bit to another according to the specific environments.

The CPF ($/ft) can be obtained using the following equation

$$CPF = \frac{\boldsymbol{Cbit}+\boldsymbol{Crig}(\boldsymbol{Tb}+\boldsymbol{Tc}+\boldsymbol{Tr})}{\boldsymbol{F}} \qquad \textbf{II.1}$$

Where:

Cbit: is the bit cost ($)

Crig: is rig cost per hour ($/h)

Tb: is the bit running time (hr)

Tc: is connection time (hr)

Tr: is drilling trip time (hr)

F: is the length of section drilled (ft).

CPF: directly affects drilling economics, but it is not dependent on actual drilling parameters.

3.1 Commercial advancement:

Consider the inverse of the price per meter, as it has just been defined:

$$\frac{1}{Pm} = \frac{M}{Po+Ph(Tm+Tf)} \qquad \textbf{II.2}$$

Each term of this equality by the price of the probe hour Ph which is constant, it comes:

$$\frac{Ph}{Pm} = \frac{M}{\frac{Po}{Ph}+(Tm+Tf)} \qquad \textbf{II.3}$$

This value is inversely proportional to the price per meter, which will be minimum when commercial advancement will be maximum. [9][10]

4 Application:

On a diagram the curve representing the cumulative advancement as a function of the rotation time on this diagram, on the negative abscissa, carry a segment OA equal to the maneuvering time and a segment AB equal to the price of the tool divided by the price per hour of the drilling rig: Po/Ph (price of the bit in probe hours). After a rotation time OT, the tool has carried out a measurement M represented by point m on the curve.

The slope of the line Bm represents the commercial advancement of the bit, we have:

Slope of

$$Bm = \frac{M}{BA+AO+OT} \qquad \textbf{II.4}$$

That we can write:

$$Bm = \frac{M}{\frac{Po}{Ph}+Tm+Tf} \qquad \textbf{II.5}$$

which is the commercial advancement of the bit. (Figure II.01)

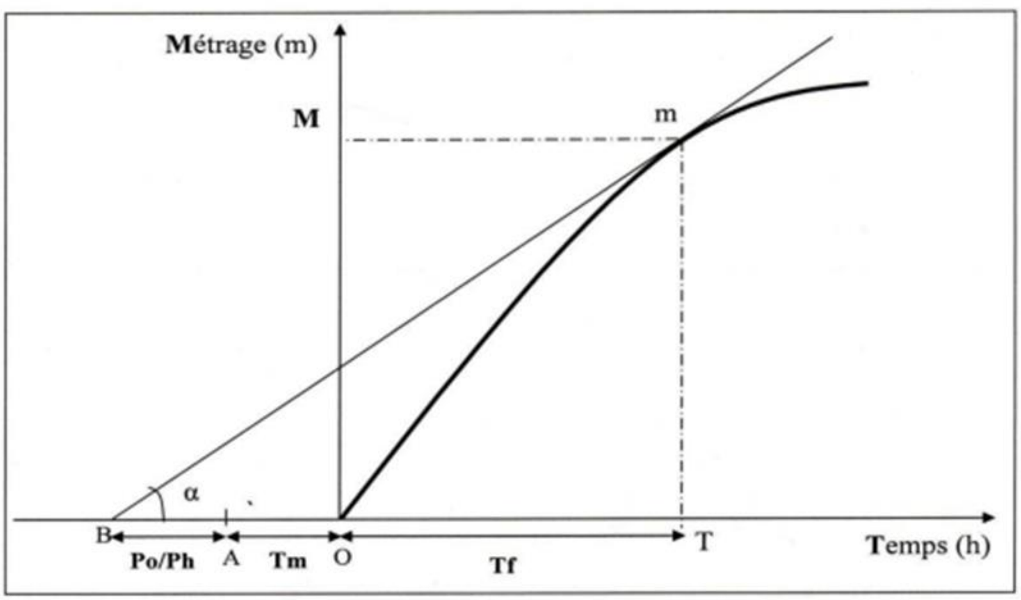


Figure II.01: The commercial advancement curves.[11]

Consequently, the price per meter will be minimum or the commercial advancement maximum when the slope of the line Bm is maximum, that is to say when this line is tangent to the advancement curve.

5 Bit selection by Break Even method:

Break-even analysis is a useful tool to study the relationship between fixed costs, variable costs and returns.

A break-even point defines when an investment will generate a positive return and can be determined graphically or with simple mathematics.

Break-even price analysis computes the price necessary at a given level of production to cover all costs. [5]

5.1 Drilling cost formula:

The most common application of drill cost formula is an evaluating the efficiency of a bit run. A large fraction of the time required to complete a well is spent either drilling or making a trip to replace the bit. The total time required to drill a given depth ***F***, can be expressed as the sum of total rotating time during the bit run ***Td***, and trip time ***Tr***.

The drilling cost formula is:

$$CPF = \frac{\boldsymbol{Cbit} + \boldsymbol{Crig}(\boldsymbol{Td} + \boldsymbol{Tr})}{\boldsymbol{F}} \qquad \textbf{II.6}$$

Were

Cbit: is the bit cost ($)

Crig: is rig cost per hour ($/h)

Td: is the drilling time (hr)

Tr: is the maneuvering time (hr)

F: is the length of section drilling (m)

This formula does not take into account certain factors which can influence the price per meter drilled (mud treatment, modification of the BHA, etc.), however it gives figures very close to the exact values.

The simplest case is to compare two tools, lowered to the same depth, in identical formations, on two neighboring wells.

The firs bit A to perform a performance, all the elements of which are known. What should be that of a second bit B so that it is more economical?

Bit B will be more profitable than tool A when: **P**mb < **P**ma

For B, the break-even point will be reached when: **P**mb = **P**ma, so:

$$CPFa = \frac{CbitB + Crig(TdB + TrB)}{Fb}$$

$$\Longleftrightarrow$$

$$Fb = \frac{CbitB + Crig(TdB + TrB)}{CPFa}$$

$$\Longleftrightarrow$$

$$Fb = \frac{Crig}{CPFa} TrB + \frac{CbitB + CrigTdB}{CPFa} \qquad \textbf{II.7}$$

This is an equation of the type Y = aX + b

- With: Y= MB: length drilled by bit B at the break-even point.
- X = Tf: drilling time of bit B at the break-even point.

$$b = \frac{CbitB + CrigTdB}{CPFa}$$

$$a = \frac{Crig}{CPFa}$$

In a Cartesian coordinate system where the drilled length (M) appears on the ordinate and the abscissa shows the

drilling time (Tf), the break-even point of one tool compared to another, is a line

To define a line, two points are sufficient:

Point X : Y = 0

$$X = {}^{-b}/_{a} = \frac{CbitB + CrigTdB}{Crig}$$

Point Y : X = 0

$$Y = b = \frac{CbitB + CrigTdB}{CPFa}$$

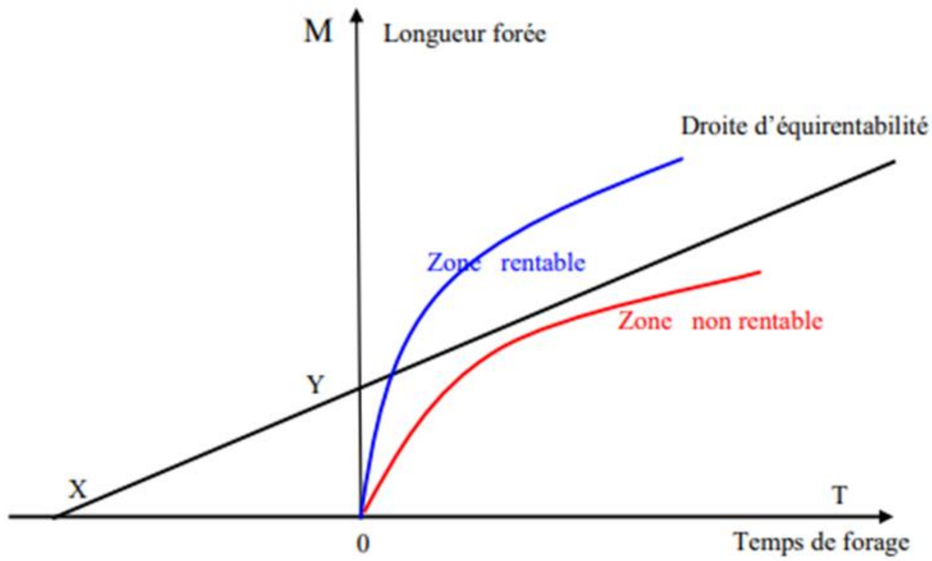


Figure II.02: the profitability curve using the Break-even method.[11]

Each point on the line (XY) represents a performance to be achieved by bit B so that its cost per meter drilled is equal to the cost per meter drilled by bit A. All the points on the graph, located above of the right (XY) represents the performances to be achieved by bit B so that its cost per meter is lower than that of bit A. Those located below the right (XY) represent all the performances for which the bit B will not be cost-effective compared to the reference bit A. (Figure II.02)

In a Cartesian reference frame showing the drilling time on the abscissa and the drilled length on the ordinate, it is necessary to draw the line which represents all of the performances that the bits to be evaluated must achieve, to be at least profitable in relation to the bit or reference bits. This line separates the plane into two portions. The upper part represents all economic performances, and the lower part represents all non-economic performances. It intersects the axes at two points X and Y.

It is possible and recommended to improve the Break-Even method by joining the graph a lithological section parallel to the ordinate axis, (Figure II.03) which can sometimes avoid a raising the bit too early.

Without the lithological cut this bit would surely have been reassembled around the sixth hour while the tool crossed the very hard B formation.

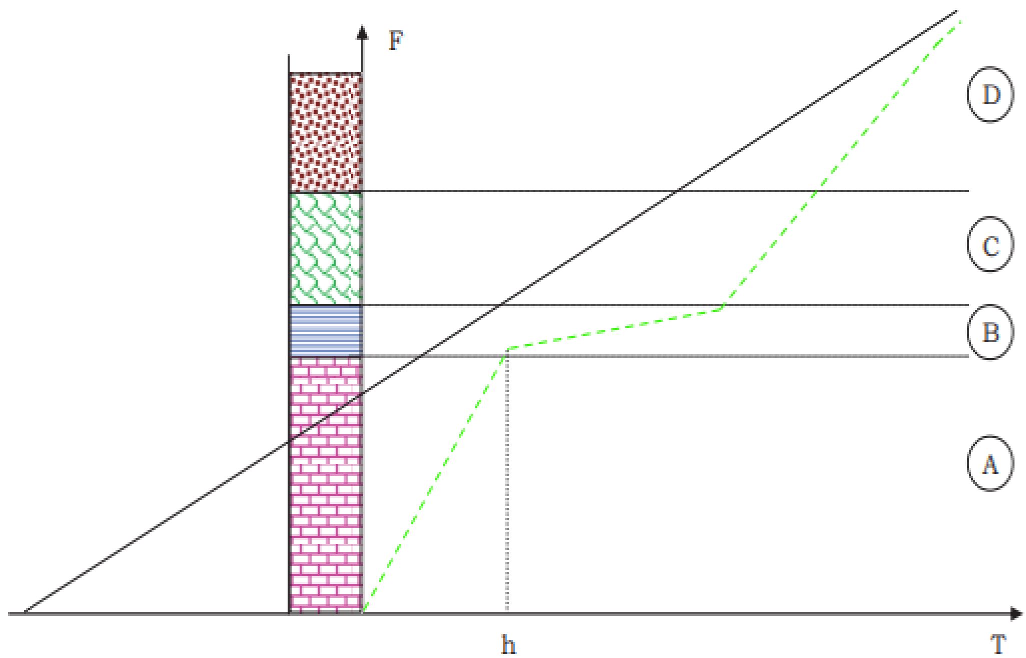


Figure II.03: Graph Break Even with a lithological section. [5]

Without the lithological section, this tool would surely have been dated around the hour h then that the bit passed the very hard B formation.

5.2 Case of turbo drilling:

Two cases can arise:

In the case where the deference run is that of turbo drilling, the procedure for calculating the price per meter is the same as previously. if the reference run is that of the tricone, we use the following formula.

$$CPF = \frac{Cbit+P+Td(Crig+Cturbine)+(Crig\times Tr)}{F} \qquad \textbf{II.8}$$

Where:

P is the load attached to the turbine (stand-by, operators, etc.)

Cturbine is the turbine cost.

The equation of the equi-profitability line becomes:

$$Fb = \frac{\boldsymbol{T}db.(\boldsymbol{C}rig+\boldsymbol{C}turbine)}{\boldsymbol{CPF}a} + \frac{(\boldsymbol{C}rig.\boldsymbol{T}rb+\boldsymbol{C}bit+\boldsymbol{P})}{\boldsymbol{CPF}a} \qquad \textbf{II.9}$$

Point X : y = 0 and x= $(Crig.Trb + Cbit + P)/(Crig + Cturbine)$

Point Y : x = 0 and $(Crig.Trb + Cbit + P)/CPFa$

CHAPTER III:

Artificial intelligence and its applications in oil field

1 Introduction:

Oil and gas industry, a cornerstone of the global economy especially in Algeria, is facing the challenge of meeting rising energy demands with dwindling resources. Traditionally, tasks like identifying drilling locations and optimizing oil recovery have been expensive, time-consuming, and often reliant on limited data. Artificial intelligence (AI) has emerged as a powerful tool to address this. AI offers a more efficient and accurate solution by automating these processes and creating expert systems that can analyze vast amounts of information. This approach not only leads to better solutions but also allows for cost savings and parallel application across different aspects of the industry.

2 Study:

This chapter looks at the application of various AI techniques and tools within the hydrocarbon industry. We will evaluate their effectiveness compared to traditional methods, quantifying both their advantages and limitations.

This chapter will delve into various AI techniques and their potential impact on the hydrocarbon sector, comparing their effectiveness to traditional methods and highlighting the potential for both qualitative and quantitative improvements.

For the past few years, the hydrocarbons business has relied heavily on the application of artificial intelligence (AI). This is because of the hydrocarbons industry's significant financial importance in the modern world, which is mostly represented by oil and gas, as well as the ongoing requirement to meet demand with the hydrocarbon reserves' finite resources.

AI seeks to develop expert systems that are capable of carrying out a wide range of tasks, from determining the best areas for drilling to putting improved oil recovery systems into place. In the past, humans have completed these jobs at tremendous expense, taking a long time, and frequently relying heavily on conjecture based on scant information. Theoretically, artificial intelligence (AI) can offer the best solution to these and other issues facing the hydrocarbons sector, frequently without requiring a thorough comprehension of the intricate underlying mechanisms that are typically governed by the laws of physics and chemistry. [12] [13]

AI can be used to do this by building process models and breaking the issue down into causal and intuition-based components. This enables a more rigorous solution that concentrates on the causative elements, which could have significant economic advantages because it can be paralleled to other sections of the industry process. AI offers and includes a range of techniques and resources for problem-solving. The utility of these solvers/types in the hydrocarbons industry will be highlighted, and a potential qualitative and quantitative evaluation of their efficacy in contrast to traditional techniques will be conducted.

Here is a paraphrased version of the introduction about using AI tools in the hydrocarbon field, along with highlighting the potential effects:

In recent years, the hydrocarbon industry, primarily represented by oil and gas, has heavily depended on artificial intelligence (AI) applications. This reliance stems from the industry's substantial economic importance globally and the ongoing need to meet demand with finite hydrocarbon reserves.

AI aims to develop expert systems capable of performing various tasks, from identifying optimal drilling locations to implementing enhanced oil recovery techniques. Traditionally, these tasks were undertaken by humans at considerable expense, time-consuming processes, and often relying heavily on speculation due to limited information. Theoretically, AI can provide optimal solutions to the challenges faced by the hydrocarbon sector without requiring a comprehensive understanding of the complex underlying mechanisms governed by physical and chemical laws.

AI achieves this by constructing process models and breaking down problems into causal and intuition-based components, enabling a more rigorous solution focused on the causative factors. This approach could yield significant economic benefits as it can be parallelized across different stages of the industry process. AI offers a range of techniques and resources for problem-solving.

The oil and gaz industry, heavily reliant on oil and gas, faces the challenge of meeting growing demand while managing finite resources. To tackle this, AI has emerged as a game-changer.

Traditionally, tasks like identifying drilling locations and optimizing oil recovery have been expensive, time-consuming, and often reliant on limited data. AI offers a more efficient and accurate solution by automating these processes and creating expert systems that can analyze vast amounts of information.

AI achieves this by building sophisticated models that break down problems into their fundamental components, focusing on the key causal factors. This approach not only leads to better solutions but also allows for cost savings and parallel application across different aspects of the industry.

This exploration will delve into various AI techniques and their potential impact on the hydrocarbon sector, comparing their effectiveness to traditional methods and highlighting the potential for both qualitative and quantitative improvements.

The oil and gas industry, a cornerstone of the global economy, is facing the challenge of meeting rising energy demands with dwindling resources. To address this, artificial intelligence (AI) has emerged as a powerful tool. AI can analyze vast amounts of data to optimize drilling locations, improve oil recovery techniques, and automate tasks that were previously time-consuming and expensive for humans. [5]

Traditional methods often relied on guesswork and limited data. AI, on the other hand, can create sophisticated models that break down complex problems into manageable components, leading to more accurate and cost-effective solutions. This approach leverages the power of physics and chemistry without requiring deep, intricate understanding.

This research delves into the application of various AI techniques and tools within the hydrocarbon industry. We will evaluate their effectiveness compared to traditional methods, quantifying both their advantages and limitations. [6]

3 Effect of using AI tool:

The potential effects of using AI tools in the hydrocarbon field:

1. Cost savings: AI can potentially reduce the substantial expenses associated with traditional human-driven processes.

2. Time efficiency: AI-based solutions could expedite tasks that typically take longer when performed by humans.

3. Improved decision-making: By leveraging data and models, AI can provide more informed and accurate decisions, reducing reliance on speculation.

4. Enhanced resource management: AI can help optimize the utilization of finite hydrocarbon reserves by identifying optimal drilling locations and implementing efficient recovery techniques.

5. Economic benefits: The rigorous, causal-based approach enabled by AI can lead to significant economic advantages through parallelization across industry processes. [7]

The introduction suggests that the application of AI tools in the hydrocarbon field can potentially enhance efficiency, decision-making, resource management, and economic benefits, while reducing costs and time requirements compared to traditional methods.

4 Artificial Intelligence:

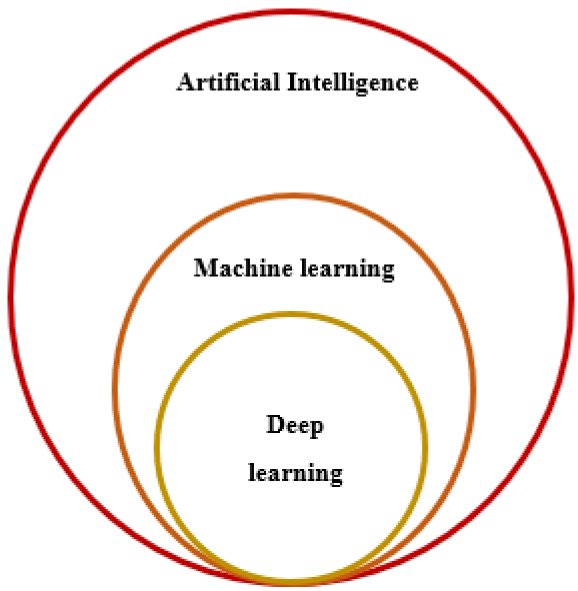


Figure III.01: Comprehensive overview of artificial intelligence (AI)

A comprehensive overview of artificial intelligence (AI):

AI is a branch of computer science that aims to create intelligent machines that can think and learn like humans. The main goals of AI include reasoning, knowledge representation, planning, natural language processing, perception, learning, and ability to move and manipulate objects.

There are several approaches to AI:

1. Rule-Based Systems use a set of rules and logic to solve problems. Expert systems are an example, encoding the knowledge of human experts.

2. Machine Learning algorithms allow systems to learn from data and make predictions/decisions without being explicitly programmed. Key techniques include neural networks, decision trees, clustering etc.

3. Deep Learning neural networks with many layers attempt to mimic the human brain in processing data in a non-linear way to recognize patterns.

4. Natural Language Processing allows machines to analyze, understand and generate human language.

5. Computer Vision deals with processing, analyzing and understanding digital images/videos.

6. Robotics combines AI software with electromechanical systems to create robots that can sense, reason and act.

Major applications of AI today include virtual assistants, predictive analytics, computer vision for self-driving cars, recommendation engines, robotics, gaming, cybersecurity and more. AI techniques like machine learning and deep learning are advancing rapidly.

However, AI also raises concerns around ethical issues like privacy, bias, transparency, and existential risk if super intelligent AI systems become uncontrollable. Regulation and governance of AI development are emerging topics.

Overall, AI has immense potential to transform many industries, while also posing challenges that need to be carefully navigated. It is a broad and rapidly evolving field. [8]

4.1 An overview of how artificial intelligence (AI) is being applied in the oil field:

The hydrocarbons industry, which includes oil and gas exploration, production, refining, and distribution, is leveraging AI technologies to optimize operations, improve safety, and reduce costs. AI can analyze massive amounts of data from sensors, drilling logs, seismic surveys and more to drive better decision making.

Upstream (Exploration and Production):

- Seismic interpretation - AI techniques like machine learning are used to automatically analyze large seismic datasets to identify potential hydrocarbon reservoirs.

- Well planning - AI systems evaluate data like formation properties to optimize well locations and drilling trajectories.

- Production optimization - AI models examine multiple variables like reservoir characteristics and surface facilities to maximize hydrocarbon recovery.

- Predictive maintenance - Machine learning detects anomalies in sensor data to predict equipment failures before they occur.

Midstream (Transportation and Storage):

- Pipeline monitoring - AI vision systems and sensor data analysis tracks pipeline integrity and detects leaks or defects.

- Terminal operations - Optimization and scheduling of storage tank activities using AI planning.

Downstream (Refining and Distribution):

- Refinery operations - AI models optimize the refining process by adjusting multiple parameters to increase efficiency.

- Supply chain - AI is used for demand forecasting, logistics route optimization, and inventory management.

- Plant safety - Computer vision and audio analytics detect safety risks at facilities using AI.

Across all segments, AI and machine learning are also being leveraged for analyzing geoscience data, automating knowledge work, mitigating cyber risks, and other key applications.

While still emerging, AI is providing the oil and gas industry with new capabilities for data-driven insights to improve productivity, profitability and sustainability. [9] [10]

As our work is based on the prediction of the ROP in order to specify the appropriate drilling bit, we concentrate in this chapter on explaining the machine learning methods.

4.2 Machine learning:

Machine learning forms a subarea within the field of artificial intelligence, enabling systems to automatically learn and improve based on experience and data exposure, without requiring explicit programming. In the context of optimizing drilling operations, machine learning techniques prove valuable for predicting the rate of penetration (ROP) and enhancing overall drilling efficiency. These methods are particularly useful when the relationship between various drilling parameters and the resulting ROP exhibits complex, non-linear characteristics that are difficult to model using traditional approaches.

By leveraging machine learning algorithms and models trained on historical drilling data, it becomes possible to capture intricate patterns and interdependencies among factors influencing ROP. This data-driven approach circumvents the need for manually programming explicit rules or equations to describe the underlying dynamics. Instead, the machine learning system can automatically infer these relationships from the data itself, enabling accurate predictions of ROP based on the specified drilling conditions and parameters. Consequently, applying machine learning in this domain facilitates more informed decision-making and operational adjustments to optimize drilling performance and efficiency. [11]

As depicted in Figure 2, machine learning techniques can be broadly classified into two main groups: supervised learning approaches and unsupervised learning approaches.

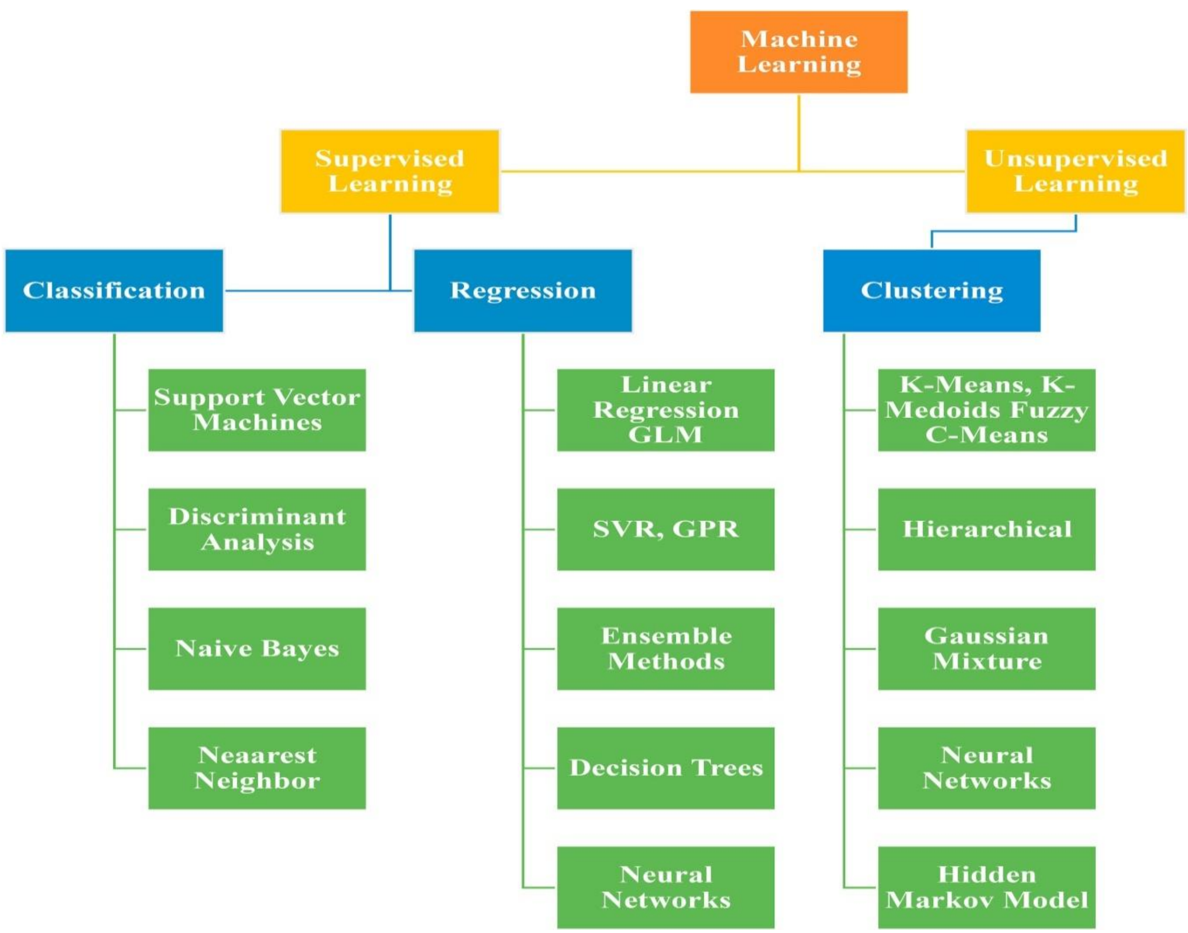


Figure III.02. Machine learning models

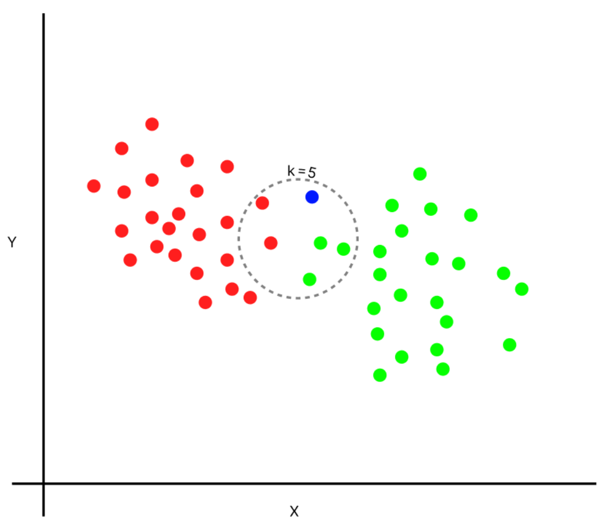
Certainly, we can explain how various machine learning regression models like KNN Regressor, XGBoost Regressor, Neural Network Regression, and Random Forest Regression can be used for predictive purposes in the hydrocarbons field:

a) KNN (K-Nearest Neighbors) Regressor:

- **How it works:** KNN Regression is a non-parametric model that predicts a value for a new data point based on the average of the values of its "k" nearest neighbors in the training dataset.
- **Advantages**: Simple, easy to understand, works well with complex datasets.
- **Disadvantages**: Can be slow for large datasets, sensitive to the choice of "k".

In the oil and gas industry, KNN regression can be used to predict continuous values like reservoir pressure, flow rates, or production volumes based on historical data and known features (e.g., well depth, porosity, etc.). KNN identifies the K closest data points to the new input in the feature space and averages their output values for prediction.

Example in Hydrocarbons: Imagine you have a dataset of well locations with corresponding production rates (oil/gas). To predict the production rate of a new well, KNN would find the "k" nearest wells (based on factors like geological features, depth, etc.) and average their production rates to give you a prediction.



FigureIII.03. K-Nearest Neighbors Regressor.

b) Random Forest Regression:

- **How it works:** A Random Forest is an ensemble method. It builds multiple decision trees, each trained on a random subset of the data and features. The predictions are then averaged across all the trees to create a final prediction.
- **Advantages:** Very accurate, works well with large datasets, relatively robust to outliers.
- **Disadvantages:** Can be less interpretable than simpler models.

This ensemble method based on multiple decision trees can be applied to predict continuous targets like well productivity, remaining useful life of equipment, or reservoir characteristics. Random forests reduce overfitting by training each tree on different subsets of data.

Example in Hydrocarbons: You can use a Random Forest to predict the optimal drilling depth for a well based on geological features, seismic data, and historical data from other wells.

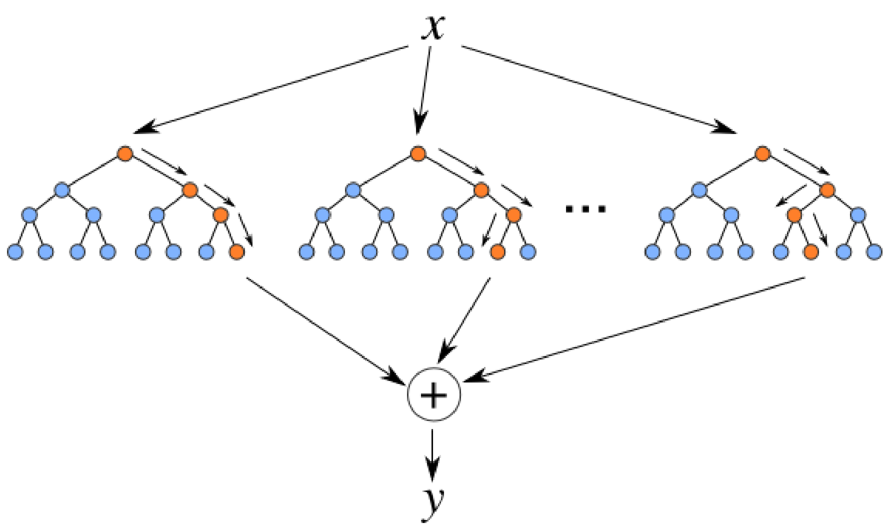


Figure. III.04 Random Forest Regressor

c) XGBoost (Extreme Gradient Boosting) Regressor:

- **How it works:** XGBoost is a powerful, tree-based ensemble method. It builds a series of decision trees, each trying to correct the errors of the previous trees. The final prediction is a weighted average of the predictions from all the trees.
- **Advantages:** Highly accurate, works well with large datasets, handles complex interactions between features.
- **Disadvantages:** Can be more complex to understand and tune compared to simpler models.

XGBoost is an advanced implementation of gradient boosted decision trees. In hydrocarbon applications, it can predict quantities like remaining oil/gas reserves, drill time, or pipeline throughput. XGBoost excels at capturing complex, non-linear relationships in data through ensemble learning.

Example in Hydrocarbons: You can use XGBoost to predict oil or gas production based on variables like well location, geological characteristics, production history, and reservoir parameters.

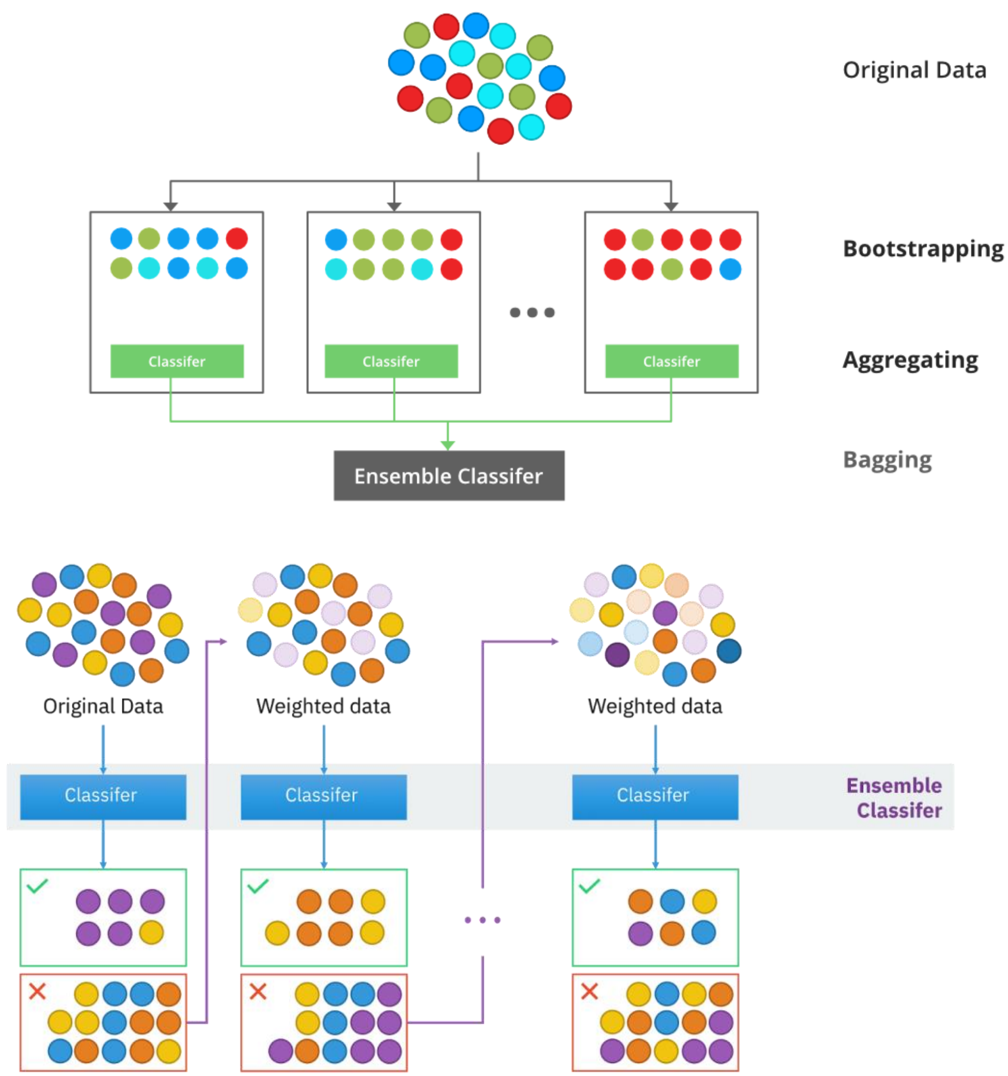


Figure. III.05 XGBoost Regressor.

d) Neural Network Regression:

- **How it works:** Neural networks are inspired by the structure of the human brain. They consist of interconnected "neurons" organized in layers. During training, the network learns weights for these connections to minimize prediction errors.
- **Advantages:** Can learn highly complex relationships, good for non-linear data.
- **Disadvantages:** Requires significant training data, can be computationally expensive, and can be harder to interpret than simpler models.

Artificial neural networks are powerful for modeling highly non-linear and complex relationships, making them suitable for tasks like forecasting future oil/gas prices or predicting drillbit wear based on drilling parameters. Different network architectures (feedforward, recurrent, etc.) can be used based on the problem.

Example in Hydrocarbons: You can train a neural network to predict the volume of oil recoverable from a reservoir based on factors like reservoir size, porosity, permeability, and geological conditions. [14]

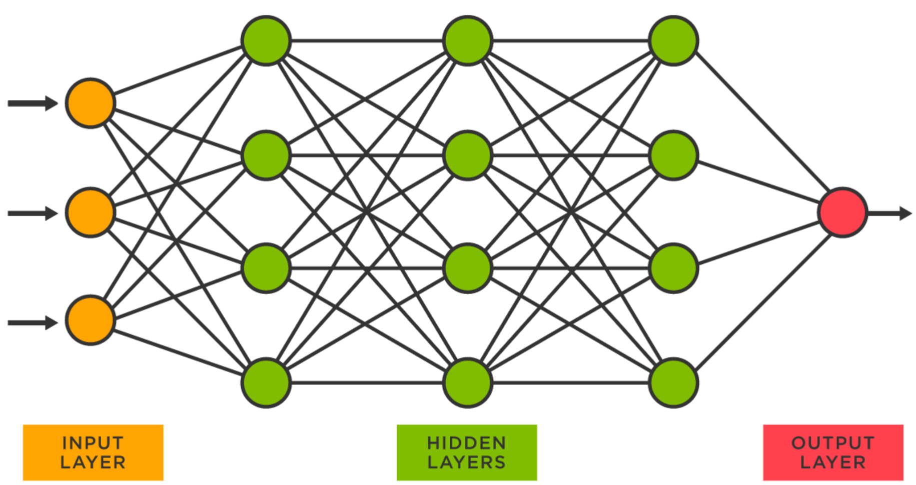


Figure. III.06 Neural Network Regressor

5 The general process:

The general process involves:

- **Data Preparation:** Historical data related to the prediction target is collected, cleaned, and transformed into suitable feature representations.
- **Model Training:** The regression model is trained on this data, automatically learning the complex mapping between input features and the continuous output variable of interest.
- **Validation and Tuning:** Models are evaluated using techniques like cross-validation. Hyperparameters are tuned to optimize performance metrics like mean squared error.
- **Prediction:** Once accurately trained, the model can make predictions on new, unseen data by applying the learned relationships to forecast future values, enable better decision making, and optimize processes.

The choice of model depends on factors like the number of input features, data volume, variable relationships, and trade-offs between model complexity, training time, and prediction accuracy required. Ensemble techniques often outperform individual models. Proper data preparation and validation are crucial for reliable predictive performance. But overall, these AI/ML techniques are enhancing data-driven decision making across the hydrocarbon value chain.

Here's an explanation of how those machine learning models work, along with examples of how they can be used in the hydrocarbons field:

1. KNN Regressor (k-Nearest Neighbors Regression):

- **How it works:** KNN Regression is a non-parametric model that predicts a value for a new data point based on the average of the values of its "k" nearest neighbors in the training dataset.
- **Example in Hydrocarbons:** Imagine you have a dataset of well locations with corresponding production rates (oil/gas). To predict the production rate of a new well, KNN would find the "k" nearest wells (based on factors like geological features, depth, etc.) and average their production rates to give you a prediction.
- **Advantages:** Simple, easy to understand, works well with complex datasets.
- **Disadvantages:** Can be slow for large datasets, sensitive to the choice of "k".

2. XGBoost Regressor (Extreme Gradient Boosting):

- **How it works:** XGBoost is a gradient boosting algorithm. It builds an ensemble of decision trees. Each tree tries to correct the errors made by the previous ones, progressively improving the model's accuracy.
- **Example in Hydrocarbons:** You can use XGBoost to predict oil or gas production based on variables like well location, geological characteristics, production history, and reservoir parameters.
- **Advantages:** Highly accurate, works well with large datasets, handles complex interactions between features.
- **Disadvantages:** Can be more complex to understand and tune compared to simpler models.

3. Neural Network Regression:

- **How it works:** Neural networks are inspired by the structure of the human brain. They consist of interconnected "neurons" organized in layers. During training, the network learns weights for these connections to minimize prediction errors.
- **Example in Hydrocarbons:** You can train a neural network to predict the volume of oil recoverable from a reservoir based on factors like reservoir size, porosity, permeability, and geological conditions.
- **Advantages:** Can learn highly complex relationships, good for non-linear data.
- **Disadvantages:** Requires significant training data, can be computationally expensive, and can be harder to interpret than simpler models.

4. Random Forest Regression:

- **How it works:** A Random Forest is an ensemble method. It builds multiple decision trees, each trained on a random subset of the data and features. The predictions are then averaged across all the trees to create a final prediction.
- **Example in Hydrocarbons:** You can use a Random Forest to predict the optimal drilling depth for a well based on geological features, seismic data, and historical data from other wells.
- **Advantages:** Very accurate, works well with large datasets, relatively robust to outliers.
- **Disadvantages:** Can be less interpretable than simpler models.

5.1 Key Steps in Using ML for Prediction:

1. **Data Collection and Preparation:** Gathering relevant data from various sources (well logs, seismic data, production data, etc.) and cleaning/preprocessing it.
2. **Feature Engineering:** Selecting the most relevant variables (features) for your model. This is crucial for accuracy.
3. **Model Selection:** Choosing the right model based on the nature of your data, complexity of relationships, and the desired accuracy.
4. **Model Training:** Training the chosen model using a labeled dataset where you have both input features and corresponding output values (e.g., production rates).
5. **Model Evaluation:** Assessing the model's performance on unseen data using metrics like mean squared error, R-squared, etc.
6. **Deployment and Monitoring:** Deploying the model to make predictions in real-world scenarios and continuously monitoring its performance.

5.2 Important Considerations:

- **Data Quality:** The accuracy of your predictions is heavily dependent on the quality and completeness of your data.
- **Model Complexity:** More complex models can lead to overfitting, where the model learns the training data too well and performs poorly on unseen data.
- **Interpretability:** Understand how your model is making predictions to ensure it's making sensible decisions.

By understanding these models and following a sound data science process, you can utilize AI to predict various aspects of the hydrocarbons industry and improve decision-making. [15]

6 AI Models:

This is more detail about how each model works and provide examples of how they can be specifically applied in the hydrocarbons industry.

1. KNN Regressor (k-Nearest Neighbors Regression):

- **How it Works:** KNN is a "lazy learner," meaning it doesn't explicitly build a model during training. Instead, it simply stores the training data. When predicting for a new data point, it finds the "k" nearest neighbors (those most similar based on features) and averages their target values.
- **Example in Hydrocarbons:**
  - **Predicting Well Production:** You have a dataset of well locations with corresponding daily production rates (barrels of oil per day). You want to predict the production rate of a newly drilled well.
  - **Steps:**
    1. **Feature Selection:** Consider features like well depth, geological formation, reservoir pressure, and distance to existing wells.
    2. **KNN:** Find the "k" (e.g., 5) nearest wells to the new well based on these features.
    3. **Prediction:** Average the production rates of those "k" nearest wells to estimate the production rate of the new well.
- **Advantages:**
  - **Simplicity:** Easy to understand and implement.
  - **Flexibility:** Works well with complex datasets and non-linear relationships.
- **Disadvantages:**
  - **Computational Cost:** Can be slow for large datasets, especially when calculating distances between all data points.
  - **Sensitivity to "k":** The choice of "k" significantly impacts the prediction. Too small a "k" leads to high variance (sensitive to outliers), while too large a "k" leads to high bias (too smooth predictions).

2. XGBoost Regressor (Extreme Gradient Boosting):

- **How it Works:** XGBoost is a powerful, tree-based ensemble method. It builds a series of decision trees, each trying to correct the errors of the previous trees. The final prediction is a weighted average of the predictions from all the trees.

- **Example in Hydrocarbons:**
  - **Predicting Reservoir Recovery:** You want to predict the total amount of oil recoverable from a newly discovered reservoir.
  - **Steps:**
    1. **Feature Engineering:** Identify relevant features like reservoir size, porosity, permeability, and geological characteristics.
    2. **XGBoost Training:** Train the XGBoost model on a dataset of known reservoirs with corresponding recovery values.
    3. **Prediction:** Input the features of the new reservoir into the trained XGBoost model to obtain a prediction of the total recoverable oil.
- **Advantages:**
  - **High Accuracy:** Generally, achieves high accuracy, especially with large datasets.
  - **Robustness:** Less sensitive to outliers and handles missing data well.
  - **Feature Importance:** Provides insights into which features are most important for prediction.
- **Disadvantages:**
  - **Complexity:** More complex to understand and tune than simpler models.
  - **Interpretability:** Can be challenging to interpret the exact decision-making process of the model.

3. Neural Network Regression:

- **How it Works:** Neural networks consist of interconnected "neurons" organized in layers. The network learns weights for these connections during training to minimize errors in predictions.
- **Example in Hydrocarbons:**
  - **Predicting Well Production Decline:** You want to predict the rate at which production will decline from a well over time.
  - **Steps:**
    1. **Feature Engineering:** Use features like well production history, reservoir pressure, and wellbore characteristics.
    2. **Neural Network Training:** Train a neural network on a dataset of wells with production decline curves.

3. **Prediction:** Input the features of a new well into the trained network to obtain a predicted decline curve.

- **Advantages:**
  - **Non-Linear Relationships:** Can model highly complex and non-linear relationships between features and target values.
  - **Feature Learning:** Can automatically learn important features from raw data.
- **Disadvantages:**
  - **Data Requirements:** Requires a significant amount of labeled training data to perform well.
  - **Computational Cost:** Training can be computationally expensive, especially for large networks.
  - **Interpretability:** Can be difficult to interpret the internal workings of the network.

4. Random Forest Regression:

- **How it Works:** Random Forest is an ensemble method that combines multiple decision trees. Each tree is trained on a random subset of the data and features, and the final prediction is the average of the predictions from all the trees.
- **Example in Hydrocarbons:**
  - **Predicting Oil Quality:** You want to predict the API gravity (a measure of oil density) of oil produced from a new well.
  - **Steps:**
    1. **Feature Selection:** Use features like geological formation, well depth, and historical oil quality data from nearby wells.
    2. **Random Forest Training:** Train a random forest on a dataset of wells with known API gravity.
    3. **Prediction:** Input the features of the new well into the trained forest to obtain a predicted API gravity.
- **Advantages:**
  - **High Accuracy:** Generally, achieves good accuracy and is robust to overfitting.
  - **Feature Importance:** Provides insights into which features are most important for prediction.

- **Handling Outliers:** Less sensitive to outliers in the data than single decision trees.

- **Disadvantages:**
  - **Interpretability:** Can be less interpretable than simpler models, as the predictions are based on a combination of many trees.

Remember: The choice of the best model will depend on the specific problem, the availability of

Data, the desired level of accuracy and interpretability.

7 Specific examples of using AI in oil field:

Let's delve deeper into each model and provide more specific examples of how they can be applied in the hydrocarbons field.

1. KNN Regressor (k-Nearest Neighbors Regression):

- **How it works:** KNN Regression, in its simplest form, finds the "k" data points in the training set that are closest to a new data point, based on a distance metric. Then it takes the average of the target values (e.g., production rates) of these "k" nearest neighbors to predict the value for the new data point.
- **Hydrocarbons Examples:**
  - **Predicting Production Rates:** Imagine you have a dataset of well locations with corresponding production rates (oil/gas). You want to predict the production rate of a new well at a specific location. KNN would:
    - Calculate the distance between the new well location and all existing well locations in the dataset.
    - Select the "k" closest wells (e.g., k=5).
    - Average the production rates of those "k" closest wells to predict the production rate of the new well.
  - **Estimating Well Decline Rates:** You can use KNN to predict how quickly a well's production will decline over time based on the decline rates of similar wells. You'd measure the similarity of wells based on factors like reservoir characteristics, well depth, and production history.

- **Advantages:**
    - **Simplicity:** KNN is a very easy algorithm to understand and implement.
    - **No Assumptions:** It makes no assumptions about the distribution of your data.
    - **Works Well with Complex Data:** It can handle complex relationships between features.
- **Disadvantages:**
    - **Computational Cost:** Can be slow for large datasets, especially when calculating distances between many data points.
    - **Sensitivity to "k":** The choice of "k" can significantly impact the model's accuracy. You need to carefully select the optimal value.
    - **Susceptible to Outliers:** KNN is sensitive to outliers in the data because they can disproportionately influence the average.

2. XGBoost Regressor (Extreme Gradient Boosting):

- **How it works:** XGBoost, in essence, builds a series of decision trees. Each tree tries to correct the errors made by the previous ones. It uses a gradient descent algorithm to minimize the prediction errors, making the model increasingly accurate.
- **Hydrocarbons Examples:**
    - **Predicting Oil/Gas Reserves:** XGBoost can predict the volume of oil or gas reserves in a reservoir based on factors like reservoir size, porosity, permeability, geological structures, and historical production data.
    - **Optimizing Well Placement:** You can use XGBoost to predict the best location for a new well by considering factors like reservoir characteristics, proximity to existing infrastructure, and potential for production.
    -
- **Advantages:**
    - **High Accuracy:** XGBoost is known for achieving very high accuracy on many datasets.
    - **Handling Large Datasets:** It is designed to handle large and complex datasets efficiently.

  - **Feature Interactions:** It handles interactions between features very well, leading to improved accuracy compared to models that only consider individual features.
- **Disadvantages:**
  - **Complexity:** XGBoost is more complex than simpler models like KNN and can be harder to understand and tune.
  - **Overfitting:** It can overfit to the training data, especially if the dataset is small or if the model is too complex.

3. Neural Network Regression:

- **How it works:** Neural networks are inspired by the structure of the human brain. They consist of interconnected "neurons" organized in layers. During training, the network learns weights for these connections to minimize prediction errors. Each neuron performs a simple calculation (activation function) and passes the result to the next layer. The process repeats until the network reaches the output layer, which provides the prediction.
- **Hydrocarbons Examples:**
  - **Predicting Production Decline Curves:** You can train a neural network to predict how the production rate of a well will decline over time, taking into account factors like reservoir characteristics, well depth, and production history.
  - **Reservoir Simulation:** Neural networks can be used to create highly complex models of reservoir behavior, which can then be used to simulate various scenarios and make decisions about production strategies.
- **Advantages:**
  - **Learning Complex Relationships:** Neural networks are excellent at learning highly complex and nonlinear relationships between features.
  - **Handling Large Datasets:** They can handle very large datasets and learn from complex patterns.
- **Disadvantages:**
  - **Data Requirements:** Neural networks require a large amount of training data to perform well.
  - **Computational Cost:** Training can be computationally expensive, requiring powerful hardware.

  - **Black Box:** It can be difficult to interpret how a neural network is making predictions, which can make it harder to trust the results.

4. Random Forest Regression:

- **How it works:** A Random Forest is an ensemble method. It builds multiple decision trees, each trained on a random subset of the data and features. The predictions are then averaged across all the trees to create a final prediction.
- **Hydrocarbons Examples:**
  - **Predicting Well Production:** A Random Forest can predict the production rate of a well based on factors like reservoir characteristics, geological formations, and well depth.
  - **Optimizing Well Completions:** You can use a Random Forest to predict the optimal completion design for a well (e.g., the number and placement of perforations) to maximize production.
- **Advantages:**
  - **High Accuracy:** Random Forests are known for achieving high accuracy, particularly on complex datasets.
  - **Robust to Outliers:** They are relatively robust to outliers in the data, making them suitable for datasets with noisy or missing values.
  - **Feature Importance:** They can provide insight into the importance of different features in making predictions.
- **Disadvantages:**
  - **Complexity:** Can be more complex to understand than simpler models.
  - **Overfitting:** Can overfit to the training data if the number of trees is too large.
  - **Computational Cost:** Can be more computationally expensive than simpler models, especially for very large datasets.

8 General Tips for Applying ML in the Hydrocarbons Field:

- **Data is King:** The quality of your data is paramount. Ensure that it is accurate, complete, and representative of the problem you are trying to solve.

- **Feature Engineering is Key:** The right features can make or break your model. Carefully consider which features are likely to be most predictive of your target variable.
- **Experiment and Iterate:** Try different models and configurations to find the best one for your specific problem.
- **Start Small:** Don't try to solve everything at once. Start with a simpler problem and build up from there.
- **Focus on Interpretability:** Strive for models that are understandable and explainable, so you can trust their predictions and make informed decisions.

By leveraging these ML techniques and following these tips, you can unlock the potential of AI to make better predictions and drive more effective decision-making in the hydrocarbons industry. [16]

9 Importance of Artificial Intelligence in the oil and Gaz Industry:

In the fast-paced world of the twenty-first century, industries all over the world are undergoing a dramatic transformation, mostly due to the unrelenting progress of technology. Among these, the oil and gas industry, which was once seen to be a stronghold of conventional methods, is now leading the way in a technological revolution. Artificial intelligence (AI), a tool that is not only altering the industry's future but also reshaping its operating framework, is a major driving force behind this revolution. AI is revolutionizing the sector in a number of areas, including environmental stewardship, predictive maintenance, and exploration and production. Through the utilization of machine learning, big data analytics, and sophisticated computing, the oil and gas industry is opening up previously unimaginable opportunities. The application of AI is guaranteeing safer, more sustainable, and economical methods in addition to improving operational efficiencies.

As we delve deeper into the transformative impact of AI in the oil and gas industry, we uncover the myriad ways in which this technology is not just an add-on but a fundamental driver of change. Whether it's through improved prediction accuracy in exploration, enhanced safety protocols, or the development of sustainable practices, AI is undeniably a cornerstone in the modernization of the oil and gas sector.

Because AI is able to solve problems in real time or on paper, it has emerged as the technology with the greatest growth rate in recent years. Moving to the hydrocarbons sector, AI works to find the best answer and offers many suggestions to produce high-quality outputs quickly.

9.1 Benefits of artificial intelligence in Hydrocarbons industry

For oil and gas companies, I could open doors to insights throughout the entire value chain. It's stimulating new efficiency in distribution, production, and exploration while assisting the industry in realizing enormous benefits. However, there are still many tasks to be completed and difficulties to overcome because the technology is still in its early phases of development and implementation. [17]

a) Increased efficiency:

There are many areas where AI continues to improve efficiency in the oil and gas industry, such as equipment maintenance and reservoir modeling. For instance, AI-powered systems using machine learning and deep learning can help in upstream operators find optimal wells.

b) Reduced costs:

The use of AI is seeing firms realize cost reduction in various ways. Operational efficiency is achieved by using AI systems, helping to reduce costs while improving production. For instance, predictive maintenance precisely and reliably identifies assets likely to fail, saving costs on scheduled maintenance and costly aftermaths.

c) Reducing Non-Productive Time:

Non-productive time, a significant cost factor in oil and gas operations, accounts for 20 to 25% of all rig operating time each year. AI applications have been developed to reduce this time significantly, thereby saving billions of dollars in lost revenue for drill rigs.

d) Streamlining operations:

AI brings tremendous benefits to the back office, helping in decision making, repetitive task management, and supply chain management and monitoring.

For example, it is possible to leverage machine learning to sift through large data sets to identify patterns and provide recommendations to assist operators with various tasks. Analysis of transportation routes can lead to road improvements, while weather and pothole data can lead to safer working conditions.

10 Applications of Artificial Intelligence in oil field:

The oil and gas industry is embracing Artificial Intelligence (AI), particularly Machine Learning (ML) methods, to enhance efficiency, optimize operations, reduce costs, and mitigate risks. Here's a breakdown of key applications and advantages:

10.1 Artificial Intelligence in the Oil and Gas Industry: A Summary of Key Applications:

Theme	Applications	Advantages	Examples	Sources
Exploration and Production	- **Reservoir detection and mapping:** Analyzing seismic images, identifying geological formations	- Improved accuracy and speed of analysis	- Machine learning algorithms to identify the most promising areas	[18,19,20]
	- **Production forecasting:** Analyzing well data, reservoir modeling	- Optimized production and cost reduction	- Predictive models to estimate the amount of oil or gas extractable	[21,22]
	- **Well management:** Optimizing production, detecting failures and anomalies	- Increased well performance and reduced downtime	- Machine learning systems to monitor operations and detect problems	[23,24]
Refining and Petrochemicals	- **Process optimization:** Real-time control of production parameters, predicting failures	- Improved efficiency, safety, and profitability	- Predictive models to identify optimal operating conditions	[25,26]
	- **Inventory management:** Optimizing inventory management, predicting demand	- Reduced storage costs and losses	- Machine learning algorithms to predict future needs and optimize inventory	[27,28]

Distribution and Marketing	- **Demand forecasting:** Analyzing consumption data, market trends	- Improved planning and supply chain management	- Predictive models to estimate future demand and optimize distribution flows	[29,30]
	- **Logistics optimization:** Planning delivery routes, managing inventory	- Reduced transportation costs and delivery times	- Optimization algorithms to plan the most efficient routes and manage inventory optimally	[31,32]

Note: This table is just a glimpse of the main applications of AI in the oil and gas industry. There are many other applications and numerous additional information sources available on the subject.

10.2 Artificial Intelligence in the Oil and Gas Industry: Detailed Applications and Advantages (with References):

a) Exploration and Production:

Application	Description	Advantages	Example	References
Seismic Data Analysis	- Interpreting seismic data to identify potential oil and gas reservoirs. - Using AI to detect patterns and anomalies that are difficult for humans to spot.	- Increased accuracy and speed of analysis. - Reduced time and cost of exploration.	- Deep learning algorithms trained on large datasets of seismic data to identify potential reservoirs with high confidence.	[33,34]
Reservoir Characterization	- Using AI to analyze well logs, core samples, and other data to create a detailed picture of the reservoir. - This	- More accurate reservoir models. - Improved understanding of the reservoir behavior.	- AI models trained to predict reservoir properties such as porosity,	[35,36]

	information is used to optimize production and recovery rates.		permeability, and fluid saturation.	
Well Planning and Optimization	- Using AI to optimize well placement, drilling parameters, and production strategies. - This results in more efficient and cost-effective well development.	- Reduced drilling costs. - Increased production rates. - Reduced environmental impact.	- AI algorithms to optimize well trajectory and completion design.	[37,38,39]
Production Forecasting	- Using AI to analyze historical production data and predict future production rates. - This information is used to plan production schedules and optimize reservoir management.	- Improved accuracy of production forecasts. - Better planning and resource allocation.	- Machine learning models trained to predict production decline curves and optimize production strategies.	[40,41]
Well Monitoring and Optimization	- Using AI to monitor well performance and detect potential problems. - This allows for early intervention to prevent production losses and safety incidents.	- Reduced downtime. - Increased well uptime. - Improved safety.	- AI systems to monitor well pressure, flow rates, and other parameters to detect anomalies and alert operators.	[42,43,44]

Drilling Automation	- Using AI to automate drilling operations, such as drilling trajectory control and mud weight management. - This improves efficiency and safety.	- Increased drilling speed. - Reduced risk of accidents.	- Autonomous drilling systems that use AI to optimize drilling parameters and minimize human intervention.	[45,46]

b) Refining and Petrochemicals:

Application	Description	Advantages	Example	References
Process Optimization	- Using AI to optimize process parameters in refineries and petrochemical plants. - This results in increased efficiency, reduced energy consumption, and improved product quality.	- Increased production efficiency. - Reduced operating costs. - Improved product quality.	- AI models to predict optimal operating conditions for different process units.	[47,48,49]
Predictive Maintenance	- Using AI to predict equipment failures and schedule maintenance proactively. - This reduces downtime, improves safety, and lowers maintenance costs.	- Reduced downtime. - Improved safety. - Reduced maintenance costs.	- Machine learning models to predict equipment failures based on sensor data and historical maintenance records.	[50,51]
Product Quality Control	- Using AI to monitor product quality and identify potential	- Improved product quality. - Reduced waste. -	- AI systems to analyze product samples and identify	[52,53]

	problems. - This ensures that products meet quality standards and reduces the risk of product defects.	Improved customer satisfaction.	deviations from quality specifications.	
Inventory Management	- Using AI to optimize inventory levels and manage supply chains. - This reduces storage costs and ensures that there is always enough product available.	- Reduced storage costs. - Improved supply chain efficiency.	- AI algorithms to optimize inventory levels based on demand forecasts and other factors.	[54,55]

c) Distribution and Marketing:

Application	Description	Advantages	Example	References
Demand Forecasting	- Using AI to predict demand for petroleum products. - This information is used to optimize production, distribution, and pricing strategies.	- Improved accuracy of demand forecasts. - Better planning and resource allocation.	- Machine learning models to predict demand for different petroleum products based on historical data and market trends.	[56,57]
Price Optimization	- Using AI to optimize pricing strategies for petroleum products. - This takes into account factors such as demand, competition, and market conditions.	- Improved profitability. - Increased market share.	- AI algorithms to dynamically adjust prices based on real-time market data and demand forecasts.	[58,59]
Logistics Optimization	- Using AI to optimize logistics operations,	- Reduced transportation	- AI systems to plan optimal transportation	[60,61]

	such as transportation routes, inventory management, and warehouse operations. - This reduces costs and improves efficiency.	costs. - Improved delivery times. - Increased efficiency.	routes, optimize inventory levels, and manage warehouse operations.	
Customer Relationship Management	- Using AI to improve customer service and build relationships with customers. - This includes providing personalized recommendations and responding to customer inquiries.	- Improved customer satisfaction. - Increased customer loyalty.	- AI chatbots to answer customer inquiries and provide personalized recommendations.	[62,63]

d) Other Applications:

Application	Description	Advantages	Example	References
Environmental Monitoring	- Using AI to monitor environmental conditions, such as air and water quality, in oil and gas operations. - This helps to ensure compliance with environmental regulations and protect the environment.	- Improved environmental performance. - Reduced environmental risks.	- AI systems to monitor air and water quality, identify potential environmental hazards, and provide early warning alerts.	[64,65]
Safety and Risk Management	- Using AI to identify potential safety hazards and manage risks in oil	- Improved safety performance. -	- AI systems to analyze historical safety data, identify	[66,67]

	and gas operations. - This improves safety and reduces the risk of accidents.	Reduced accidents.	potential hazards, and recommend safety measures.	
Data Analytics	- Using AI to analyze large datasets of oil and gas data, such as production data, financial data, and environmental data. - This provides valuable insights that can be used to improve decision-making.	- Improved decision-making. - Increased efficiency.	- AI tools to analyze data, identify trends, and provide insights to support decision-making.	[68,69]
Cybersecurity	- Using AI to protect oil and gas infrastructure from cyberattacks. - This includes detecting malicious activity and preventing unauthorized access to critical systems.	- Improved cybersecurity. - Reduced risk of cyberattacks.	- AI systems to detect and prevent cyberattacks, identify vulnerabilities, and monitor network traffic.	[70,71]

AI, particularly ML methods, is transforming the oil and gas industry, offering significant benefits across the value chain. By embracing AI, the industry can optimize operations, reduce costs, improve safety, and enhance environmental performance. However, it is crucial to address the challenges associated with AI implementation to maximize its potential and ensure its responsible deployment.

11 Challenges and Limitations of AI in Hydrocarbons:

The limitations of artificial intelligence (AI) encompass various aspects that pose challenges to its development, widespread adoption and application. Some of the key challenges and limitations include:

11.1 Data Availability and Quality:

AI algorithms heavily rely on large volumes of high-quality data for training and decision-making. However, acquiring and maintaining such data can be a daunting task. Issues such as data scarcity, incompleteness, and inconsistency can hinder the performance of AI systems. Additionally, concerns regarding data privacy and security arise when handling sensitive information. Striking a delicate balance between data utility and privacy is essential to build trust and comply with regulations, while still extracting valuable insights from data.

11.2 Interpretability and Explainability:

Many AI models, particularly deep learning algorithms, operate as "black boxes." They provide outputs without clear explanations or insights into their decision-making process. This lack of interpretability and explainability can be problematic, particularly in critical applications/domains such as healthcare and finance where transparency and accountability are key. Researchers are actively exploring methods to enhance interpretability and provide explanations for AI model outputs, aiming to build trust and facilitate human-AI collaboration.

11.3 Lack of Emotional Intelligence and Human Intuition:

Human decision-making often involves intuition, gut feelings, and instinctive judgments based on subtle cues or prior experiences. AI systems rely on data and algorithms, lacking the intuitive leaps that humans can make when faced with ambiguous or uncertain situations. This limitation hinders its ability to understand and respond to complex human emotions, interpersonal dynamics, and subjective aspects of decision-making that require empathy and intuition. [72]

12 Conclusion:

Based on the investigation of the application of AI in oilfield development, it could be concluded that the intelligent oilfield is on its way towards integration of business application, coordination of decision and deployment, real-time production management, visualization of comprehensive research and sharing of information resources. In the end, AI oilfield will merge exploration, development, gathering, refining, and management, among other processes, to form an intelligent ecosystem. Collaboration across all levels, geographies, and disciplines could be realized based on the ecosystem in order to prolong the life cycle of the oilfield, enhance the effectiveness and caliber of decision-making, lower costs and boost economic benefit, and ultimately complete the shift from digital to artificial intelligence (AI).

CHAPTER IV:

Computational prediction of drill penetration rate (ROP) and optimization of drill bit selection

1 Introduction:

The rate of penetration (ROP) prediction is crucial for optimizing drilling operations, reducing costs, and improving overall drilling efficiency. Traditional methods for ROP prediction often rely on empirical correlations and expert knowledge, which can be limited by data availability and the complexity of drilling processes.

This chapter explores the potential of machine learning techniques for enhancing ROP prediction capabilities. We investigate four commonly used models: K-Nearest Neighbors (KNN), Random Forest, XGBoost, and Artificial Neural Networks (ANN). These models offer a diverse range of approaches for capturing complex relationships between input variables and the target variable, ROP.

The objective is to compare the predictive performance of these models and assess the impact of hyperparameter tuning and cross-validation on their accuracy. By analyzing model performance on two distinct wells, we aim to identify the most promising technique for ROP prediction and provide insights for future research and development.

The drill bit selection process is intricately tied to accurate ROP predictions. Drilling engineers can improve drilling performance and efficiency by using the best predictive model to help them decide which drill bit is best for a given set of geological circumstances. The ramifications of machine learning-driven ROP predictions for drill bit selection are explored in this chapter. By contrasting the four models, we want to find the best technique for ROP prediction as well as a methodical approach to drill bit selection that maximizes drilling efficiency while minimizing operating expenses. Future developments in drilling equipment and technique can be guided by the insights gathered from this study, guaranteeing a more data-driven and effective approach to well design and execution.

2 Study area:

2.1 Area:

The Rhourde Nouss region contains six major hydrocarbon producing fields. These are the Rhourde Nouss Central, Rhourde Nouss Southwest, Rhourde Nouss Southeast, Rhourde Adra, Rhourde Chouff, and Rhourde Hamra fields. For completeness, other small hydrocarbon accumulations such as Rhourde Nouss Northeast, Meksem, and Dra Allal are included in this report. Reference is also made to the Hamra field to the south for extension of the geologic trends from that area. The term “Rhourde Nouss Complex” is used when referring specifically to the four central fields: Rhourde Nouss Central, Rhourde Nouss Southwest, Rhourde Nouss Southeast, and Rhourde Adra. They are grouped together because of their proximity and common development. There are numerous productive reservoirs in the Rhourde Nouss region. They may be arranged vertically into three major groups: the Triassic Argileux Greseux Superieur (TAGS), the Infra-TAGS, and the Ordovician. The TAGS is subdivided into a basal sand and main reservoir. The main reservoir may be arbitrarily subdivided into upper and lower, or braided stream and meandering stream, sections. For reservoir modeling, the total TAGS has been divided into 15 layers. The Infra-TAGS is first divided by the Hercynian unconformity into Triassic and Silurian reservoirs. Above the unconformity are the Triassic Intermediate (TINT I and TINT II), and the Triassic Argileux Greseux Inferior (TAGI). Below are the Silurian B2, B1, A2, and A1.

2.2 Wells situation:

- **Well 1 situation**

Hole RHAS-1 is located at the bottom of the Rhourde Nouss North East structural axis, at 3.07km north-east of the RNNE-2 well and 6.84km south-east of the EAD-1 well. It is located at the the intersection of the seismic lines In line:1453 and X line: 6813 The coordinates and altitude of the well platform are as follows

Table.IV.01: The coordinates and altitude of the well platform

UTM	Geographic	Altitudes
X = 291 861.025m	X = 291 861.025m Longitude	Zsol = 272.00m
Y = 3 300 061.001m	Latitude: 29° 48’ 56.837’’ N	Ztable = 281.00m

●Well 2 sitaution

The RHAS QH-1 well was installed below the culmination of the Rhourde Hamra Sud trend, which lies between two anticlinal axes running north-east-south-west (Rhourde Hamra axis and Rhourde Hamra Sud axis).

The RHAS QH-1 drilling site was chosen at the intersection of the 3D seismic sections (Inline 1882 and Cross-line 6713), on the basis of the structural map produced at the top of the Hamra Quartzites, the latter being established on the basis of the seismic attribute cubes, variance, dip, dip azimuth, at the surface of the seismic horizons interpreted at the top of the Hamra Quartzites.

Table.IV.02: The coordinates and altitude of the well platform

UTM	Geographic	Altitudes
X : 294 426.048 m	Longitude : 06° 52’ 16.16572’’ E	Zsol : 275 m
Y : 3 308 482.934 m	Latitude : 29° 53’ 31.83297’’ N	Ztab : 284 m

2.3 Wells description:

●Well 1 description

The table below shows several characteristics and information of RHAS-1

Table.IV.03: characteristics and information of RHAS-1

	RHAS-1	
Field	HMD	
Well classification	Exploration	
Operator	Sonatrach	
Drilling contractor	Enafor	
Drilling Rig	Enafor #13	
Surface Location	LSA Latitude Longitude	X= 291 861.025m Y= 3 300 061.001m 29° 48’ 56.837’’ N 06° 50’ 46.471’’ E
Well TD	4048m	4041m

●Well 2 description

The table below shows several characteristics and information of RHAS QH-1

Table.IV.04: characteristics and information of RHAS QS-1

Well name	Rhourde Hamra south QH-1 ST1
Code	RHAS QH-1 ST1
Country	Algeria
Basin	Berkine West
Permits	Rhourde Nouss - Ain Amedjane
Block	246
Type of survey	Exploration
UTM coordinates	X : 294 426.048 m Y : 3 308 482.934 m
Geographical coordinates	Longitude : 06° 52’ 16.16572’’ E Latitude : 29° 53’ 31.83297’’ N
Altitudes	Zsol : 275m Z tab : 284
Primary objectives	TAGS, Carbonate Triassic, TAGI, Silurian F6 (units: M2, M1, M0 and N1) and Hamra Quartzite Reservoir.
Drilling Rig	ENF-13
Start of drilling	18/03/2015
End of drilling	22/09/2015
TD Driller	4840 m
TD Logger	TD not reached

3 Problem formulation:

In the hydrocarbon exploration and production industry, drilling operations represent a substantial portion of the total project cost. One of the critical factors influencing drilling efficiency is the rate of penetration (ROP), which measures how quickly the drill bit can penetrate the rock formation. A higher ROP generally translates to reduced drilling time and lower overall costs. However, predicting ROP accurately is challenging due to the complex interplay of various factors, including:

- Geological factors: Rock type, formation hardness, and compressive strength.
- Drilling parameters: Weight on bit (WOB), rotary speed (RPM), and hydraulic horsepower.
- Bit design: Bit type, cutter size, and placement.
- Drilling fluid properties: Mud weight, viscosity, and flow rate.

Additionally, selecting the most appropriate drilling bit for a given formation is crucial for maximizing ROP. An incorrect bit choice can lead to premature bit wear, reduced ROP, or even catastrophic bit failure, resulting in costly tripping operations and lost time.

The drilling process in hydrocarbon exploration is complex and expensive, heavily reliant on accurate predictions of penetration rate and appropriate bit selection. Traditional methods for ROP prediction and bit selection often rely on empirical models or rules of thumb, which may not capture the full complexity of the problem or adapt well to varying geological conditions. This leads to suboptimal drilling performance and increased costs. Current methods rely on empirical data and expert judgment, often leading to inefficient drilling and costly downhole complications. This problem formulation proposes leveraging machine learning to develop a predictive model that can improve drilling performance by accurately forecasting penetration rate ROP and suggesting optimal bit selection.

3.1 Problem Statement:

Given (Inputs): A dataset, extracted from two wells, containing historical drilling data, including geological formations, drilling parameters (e.g., weight on bit, rotational speed, mud properties, depth, pressure, torque, etc.), and drilling performance metrics (e.g., penetration rate, bit wear).

Goal: Develop a machine learning model that can:

- ***Predict the penetration rate*** for a given set of drilling parameters and geological conditions.
- ***Recommend the most appropriate drilling bit*** based on the predicted penetration rate and geological characteristics.

3.2 Objectives:

The primary objectives of this project are:

- To Investigate and compare various machine learning algorithms, such as K nears-neighbors, Random Forests, Extreme Gradient Boosting machines, and neural networks, to build predictive models for ROP. The goal is to identify the most accurate and robust model that can generalize well across different geological settings.
- To identify the key factors influencing the rate of penetration and their relative importance.
- To develop a decision-support system for selecting the most appropriate drilling bit (based on geological and operational parameters.) for given drilling conditions.
- To improve overall drilling efficiency by minimizing downtime and optimizing bit selection.
- To validate the proposed models using real-world data from hydrocarbon drilling operations.
- To quantify the potential economic impact of the framework in terms of reduced drilling time, cost savings, and improved project economics.

3.3 Hypothesis:

- ➢ Machine learning techniques can effectively predict the rate of penetration with a high degree of accuracy using readily available drilling data and geological information.
- ➢ The inclusion of real-time drilling data improves the performance of ROP prediction models.
- ➢ A machine learning-based decision-support system can enhance the selection of drilling bits, leading to better drilling performance and reduced costs.

4 Methodology Overview:

The methodology involves the following steps:

- ✓ ***Data Collection:*** Gathering historical drilling data, geological information, and real-time sensor readings, including ROP, bit type, geological parameters, and operational settings.
- ✓ ***Data Preprocessing:*** Cleaning and preprocessing the data to handle missing values, outliers, and normalization to ensure its accuracy and consistency.
- ✓ ***Feature Selection:*** Selecting relevant features that influence ROP and bit performance. Applying feature scaling and transformation techniques to optimize the performance of ML models.
- ✓ ***Model Development and evaluation:*** Training and testing various machine learning models to predict ROP. Selecting the most suitable model based on performance metrics and model interpretability.
- ✓ ***Bit Selection Algorithm:*** Developing a machine learning-based algorithm that utilizes the trained model's predictions to recommend the optimal drilling bit for specific drilling conditions.
- ✓ ***Model Validation:*** Validating the models using a separate dataset and evaluating their performance using appropriate metrics.

4.1 Expected Contributions:

This study is expected to provide valuable insights and tangible outcomes, including:

- ➢ A robust machine learning model for accurate ROP prediction.
- ➢ An intelligent decision-support system for optimal drilling bit selection based on predicted ROP and other drilling conditions.
- ➢ Insights into the key factors influencing drilling performance, aiding in better planning and execution of drilling operations.
- ➢ Enhanced drilling efficiency and reduced operational costs.
- ➢ Improved safety and reduced drilling risks.

By leveraging machine learning techniques, this work aims to enhance the efficiency and effectiveness of drilling operations in the hydrocarbon industry. Accurate ROP predictions and optimal bit selection can lead to significant cost savings, reduced drilling times, and improved overall productivity. The findings of this study have the potential to drive innovation and technological advancement in drilling practices, ultimately contributing to the sustainable and profitable extraction of hydrocarbon resources

4.2 Methodology:

In this section, we will set out in detail the proposed approaches that we will take to achieve the objectives of the study.

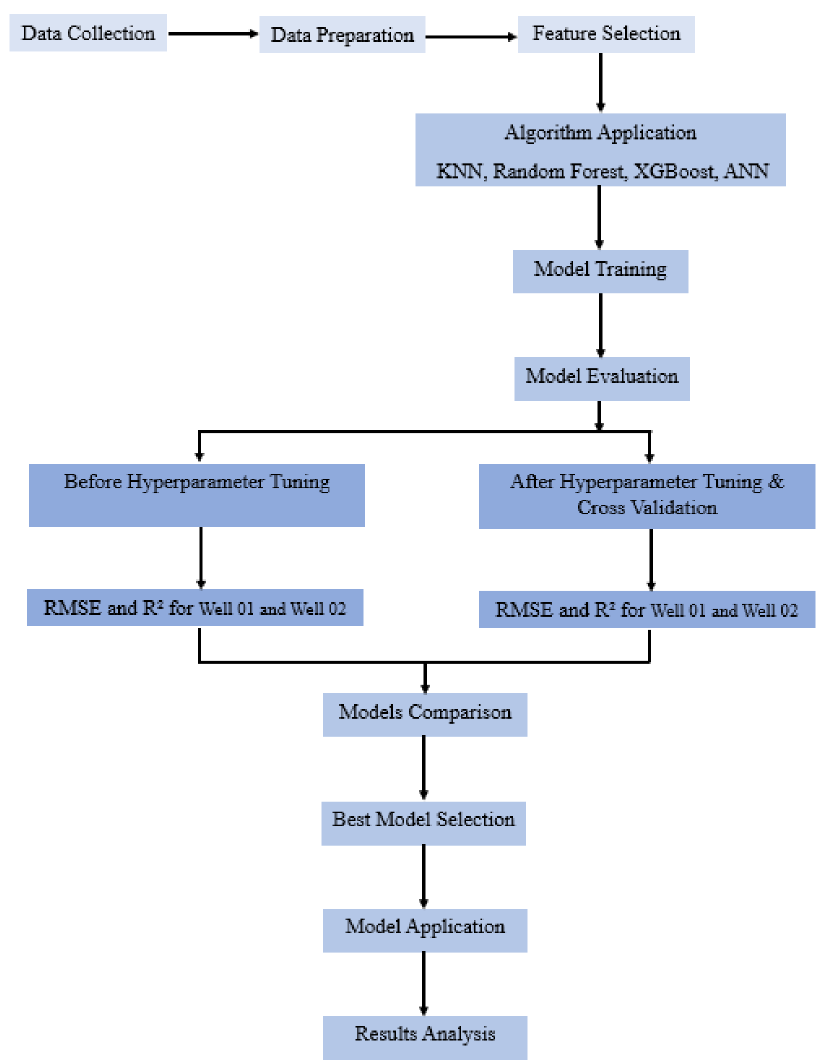


Figure IV.01: Flow chart of the proposed approaches.

5 Development Environment:

The machine learning models were developed and evaluated in Python using the Anaconda distribution, leveraging libraries such as scikit-learn for machine learning algorithms, pandas for data manipulation, NumPy for numerical computation, and matplotlib for data visualization. Jupiter Notebooks provided an interactive environment for data manipulation, model training, and visualization, enabling iterative development and code sharing. This environment facilitated the efficient exploration and implementation of various machine learning algorithms, ultimately contributing to the robust development and evaluation of the ROP prediction and bit selection system.

5.1 Data collection:

One of the most important steps in creating machine learning-based models is providing relevant data; As a result, providing higher quality data with more samples can enhance the accuracy of these estimators.

The necessary data was collected through reports submitted by Sonatrach, as the studied wells are located in the Hassi Messaoud region (southern Algeria).

The data is presented in the form of a compressed file in (RAR) format that contains an Excel file entitled (Drilling_Data) that contains a follow-up of the drilling process for each meter, in addition to two PDF files, each of which contains a comprehensive description of the well and its location, the drilling process for all geological, mechanical and hydraulic settings, and a description of all the equipment used. As well as the operations that took place during drilling.

The study was allocated to include only the 16-inch technical layer, whose depth generally ranges from (470 m to 2500 m).

Through the reports, 6 main inputs were taken, namely TVD, WOB, RPM, SPP, FLOW Pump, and TORQ, as they are the most important drilling data present in the reports and have a significant impact on the drilling process.

Also, a single target was taken which is the ROP.

Table.IV.05: data description

Feature Name	Data Type	Description	Range/Values	Units	Source
Formation_Type	Categorical	Type of geological formation	[Shale, Sandstone, Limestone, ...]	N/A	Geological Survey
Depth	Numerical (Float)	Depth of measurement	1000 - 15000	Ft	Drilling Log
ROP	Numerical (Float)	Rate of Penetration	1 - 200	ft/hr	Real-time Sensor
WOB	Numerical (Float)	Weight on Bit	5 - 50	Klbs	Real-time Sensor
RPM	Numerical (Integer)	Rotary Speed	50 - 250	Rpm	Real-time Sensor
Mud_Flow_Rate	Numerical (Float)	Drilling Fluid Flow Rate	100 - 1000	gal/min	Real-time Sensor
UCS	Numerical (Float)	Unconfined Compressive Strength	1000 - 30000	Psi	Lab Tests
Bit_Type	Categorical	Type of Drilling Bit	[PDC, Roller Cone, Hybrid, ...]	N/A	Bit Manufacturer
Cutter_Size	Numerical (Float)	Size of PDC Cutters	8 - 19	Mm	Bit Specification

Table.IV.06: A summary of real data for well 2

	Depth	WOB	RPM	TRQ	FlowIn	SPP	ROPH
count	2137	2137	2137	2137	2137	2137	2137
mean	1672	16.05	156	10236.4	3012.48	2340.05	36.87
std	617.04	4.61	16.18	2071.06	170.06	292.14	31.47
min	604	1	80	3334	1798	1337	3.4
25%	1138	13	154	8644	3012	2143	15.7
50%	1672	15	161	10096	3045	2447	26.7
75%	2206	20	166	11746	3083	2544	44.5
max	2740	27	179	15897	3167	2837	164.6

5.2 Data Preparation:

Data preparation is a critical phase in any machine learning project. It involves transforming raw data into a format that machine learning algorithms can effectively use. Here's what data preparation typically involves in our work:

a) Data Cleaning:

Data cleaning is a crucial step to ensure the quality and integrity of the dataset. It involves identifying, correcting, or removing errors, inconsistencies, and irrelevant information in the data.

● *Identify and remove duplicates*: Find and eliminate duplicate data entries, which can skew/distort analysis.

● *Correct inconsistencies:* Address typos, formatting errors, and data values that don't fit expected patterns (e.g., date formats, units).

● *Handle outliers:* Identify and manage extreme values that might be due to errors or represent legitimate anomalies. Decide whether to remove them, transform them (e.g., using mean or log transformations), or leave them depending on the context. Figure IV.02: shows an example of using a Box plot to visualize outliers in the used dataset.

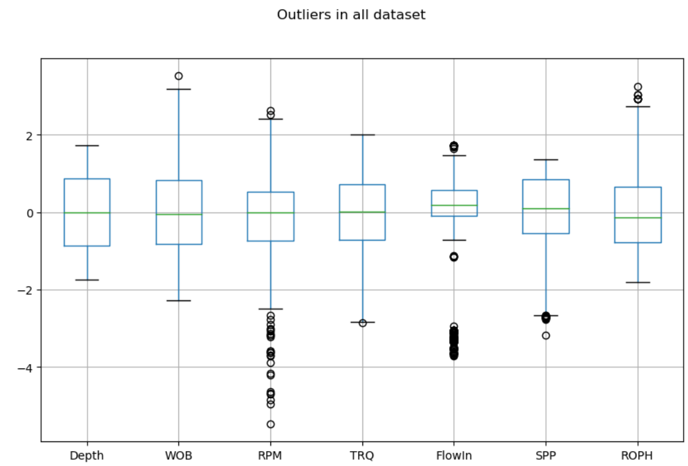


Figure IV.02: Example of Box plot for detecting outliers

b) Handling Missing Values:

Missing values can significantly affect the performance of machine learning models. It's essential to handle them appropriately. The purpose of this step is to delete, replace, or impute missing data points to maintain the integrity of the dataset.

c) Data Normalization and Scaling:

Normalization ensures that all numerical features contribute equally to the analysis and model training by bringing them to a common scale. In our work, we used the ***Min-Max Scaling*** method that transforms the values to a range between 0 and 1 (eq. 01).

$$X_{scaled} = G\,\frac{X - X_{min}}{X_{max} - X_{min}} \qquad \textbf{IV.1}$$

Table.IV.07: A summary of normalized data for well 2

	Depth	WOB	RPM	TRQ	FlowIn	SPP
Count	2137	2137	2137	2137	2137	2137
Mean	0.5	0.58	0.77	0.55	0.89	0.67
Std	0.29	0.18	0.16	0.16	0.12	0.19
Min	0	0	0	0	0	0
25%	0.25	0.46	0.75	0.42	0.89	0.54
50%	0.5	0.54	0.82	0.54	0.91	0.74
75%	0.75	0.73	0.87	0.67	0.94	0.8
Max	1	1	1	1	1	1

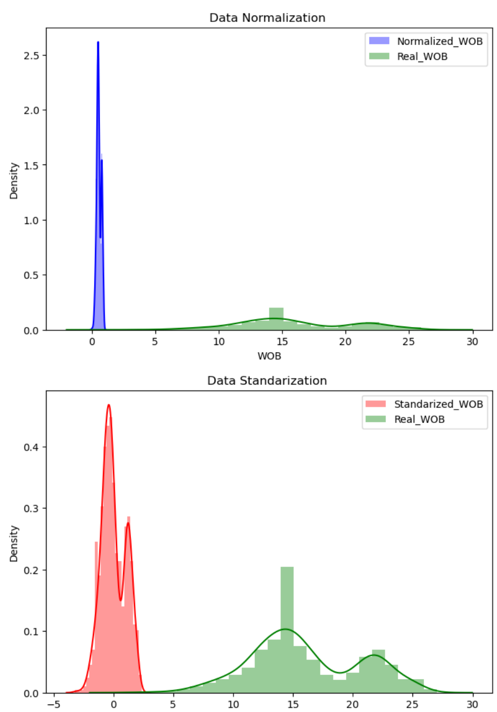


Figure IV.03: Example of data normalization

d) Exploratory Data Analysis:

This analysis is used to understand data distributions, correlations, and patterns.

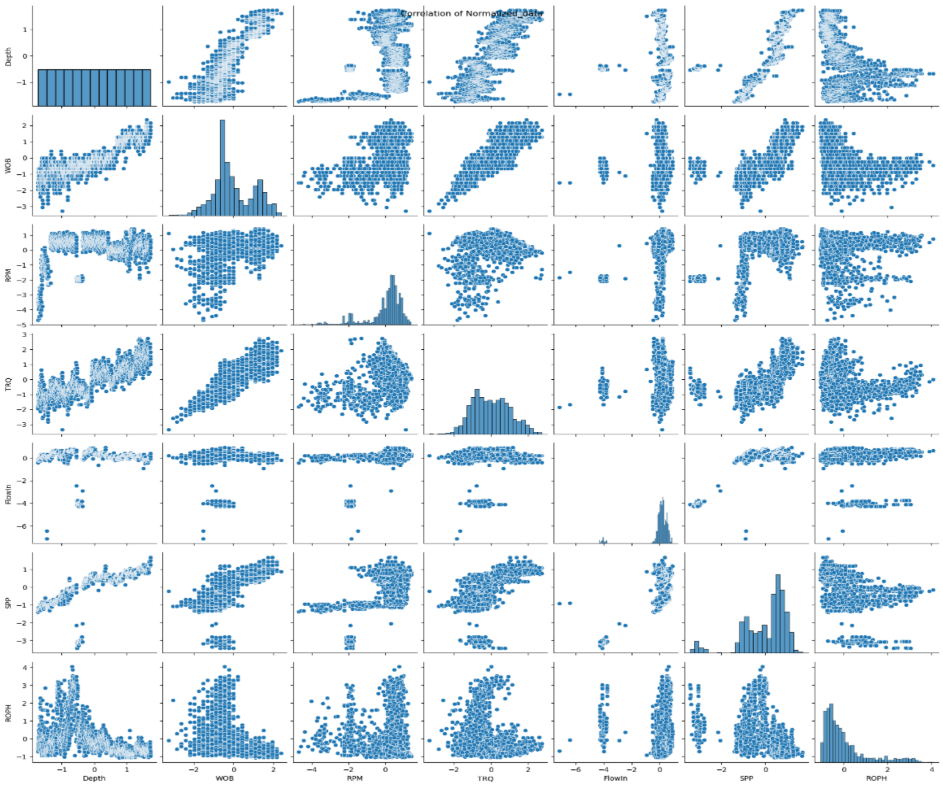


Figure IV.04: Example of data correlation (well 2)

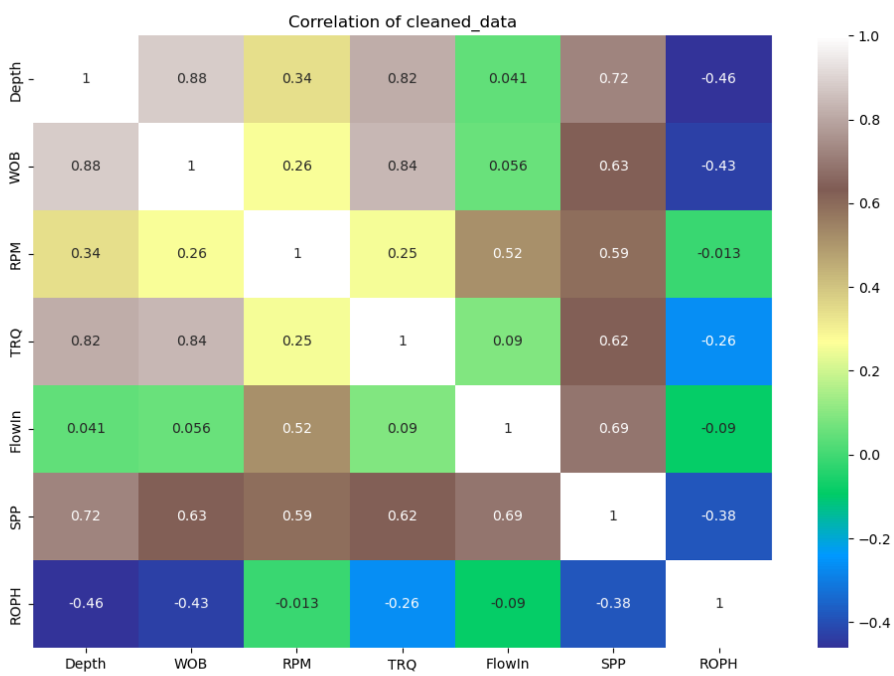


Figure IV.05: Correlation matrix of well 2 values

e) Data Splitting:

Splitting data means dividing this data into training, validation, and test sets:

► *Training*: to fit model parameters.

► *Validation*: to tune hyperparameters and prevent overfitting.

► *Testing*: to evaluate final model performance.

In our work, we first used the simple split method (70-15-15) and then the cross-validation method to ensure that the splits represent the diversity of your data (stratified sampling).

5.3 Model Development:

The goal of this study is to develop and compare various machine learning models for predicting the rate of penetration (ROP) in hydrocarbon drilling operations. Accurate ROP predictions can significantly enhance drilling efficiency and reduce operational costs. This section outlines the development and training of four different models: K-Nearest Neighbors (KNN), Random Forest, XGBoost, and Artificial Neural Networks (ANN).

a) Models' selection:

This research aims to evaluate the effectiveness of various machine-learning techniques for ROP prediction. The chosen models offer a diverse range of approaches to capturing complex relationships between input variables and the target variable.

- *K-Nearest Neighbors (KNN):* This non-parametric model classifies new data points based on their proximity to existing data points in the feature space. KNN is suitable for handling non-linear relationships and can be effective for ROP prediction due to its ability to identify similar drilling conditions and predict their corresponding performance.

- *Random Forest:* This ensemble learning method combines multiple decision trees to improve prediction accuracy and reduce overfitting. Random Forest excels at handling high-dimensional datasets and can capture complex interactions between variables, making it a suitable choice for ROP prediction.

- *Extreme Gradient Boosting (XGBoost):* This gradient boosting algorithm iteratively builds an ensemble of decision trees, weighing the importance of each tree based on its predictive performance.

XGBoost is known for its high accuracy and robustness, especially in complex datasets with numerous features.

- *Artificial Neural Networks (ANN):* This powerful model is inspired by the structure of the human brain and can learn complex non-linear relationships between input and output variables. ANNs have shown promising results in various prediction tasks and could potentially offer superior accuracy compared to other models.

b) Model Training and Evaluation:

The preprocessed dataset was split into training and testing sets, with 70% used for training and 30% for testing. Each model was trained on the training set and subsequently evaluated on the testing set. Performance was measured using the metrics: *Root Mean Squared Error* (RMSE) and *R-squared* (R^2) which are explained in Chapter III.

The suggested machine learning algorithms necessitate user-defined initialization of specific parameters (as outlined in Table.IV.07:). Within this framework, it becomes beneficial to devise an experimental design or propose methodologies for identifying the optimal parameter combinations automatically. Our approach involves exploring various techniques, including hyperparameter selection methods, to determine the most suitable parameters for each algorithm. Subsequently, we aim to compare the outcomes yielded by the default configuration of each algorithm against those achieved through our proposed parameter selection methodology.

Table.IV.08: Initialization parameters for models.

Model	Parameters	Optimal value
KNN	n_neighbors == The number of neighbors in the voting process	9
Random forests	N_estimators == The number of trees in the ensemble Max_depth == The maximum depth of each tree	100 20
XGBoosting	N_estimators == The number of trees in the ensemble Max_depth == The maximum depth of each tree Learning_rate == controls the step size at each iteration	200 20 0.1
ANN	Batch_Size== The number of training examples (in our case, drilling data points) used in one iteration of model training. Epochs== One epoch is a complete pass through your entire training dataset	10 100

To determine the optimal value of each model's parameters, we performed a ***grid search*** method with ***cross-validation.***

6 Results and Discussion:

6.1 ROP prediction:

This section presents the performance evaluation of four machine learning models - KNN, Random Forest, XGBoost, and ANN - for predicting the rate of penetration (ROP). The models were evaluated on two wells (Well 01 and Well 02) using two experimental setups: 1) a simple split without hyperparameter tuning and 2) a combination of hyperparameter tuning and cross-validation. The performance metrics used for comparison include Root Mean Squared Error (RMSE) and R-squared (R^2). The results, with and without hyperparameter tuning, are presented in Table.IV.09.

Table.IV.09: Models' performance metrics.

Model	Evaluation metrics	Without hyper-parameters tuning Simple split		With hyper-parameters tuning and Cross Validation	
		Well 01	Well 02	Well 01	Well 02
KNN	RMSE	12.789	18.730	12.551	19.220
	R^2	76.352	65.576	77.222	63.752
Random Forest	RMSE	11.675	13.973	11.697	14.125
	R^2	80.291	80.839	80.217	80.422
XGBoost	RMSE	12.297	14.111	12.056	13.232
	R^2	78.136	80.460	78.986	82.818
ANN	RMSE	12.810	18.785	12.593	18.697
	R^2	76.275	65.373	77.073	65.697

Table.IV.09 presents the performance metrics of the four models used across both wells, with and without hyperparameter tuning and cross-validation.

- **Performance Without Hyperparameter Tuning**

The results of the simple split evaluation without hyperparameter tuning show that Random Forest consistently outperforms other models on both wells, achieving the lowest RMSE and highest R^2 values. Notably, KNN and ANN exhibit similar performance, with lower R^2 values compared to Random Forest and XGBoost. XGBoost demonstrates competitive performance on Well 02, achieving an R^2 value comparable to Random Forest. However, on Well 01, its performance falls slightly behind Random Forest.

- **Performance with Hyperparameter Tuning and Cross-Validation**

After incorporating hyperparameter tuning and cross-validation, the performance of the models improved significantly. While the overall trends observed in the simple split experiment remain consistent, the performance differences between models become more pronounced.

  - *Random Forest*: This model continues to demonstrate remarkable consistency, achieving minimal improvement in performance after hyperparameter tuning. Its stability across both wells and scenarios highlights its robustness and suitability for ROP prediction.
  - *XGBoost:* XGBoost shows significant improvement with hyperparameter tuning, particularly on Well 02, where its R^2 value surpasses the others by a substantial margin. This suggests that XGBoost benefits greatly from careful optimization of its hyperparameters, potentially due to its ability to learn complex relationships between input variables.
  - *KNN:* Despite hyperparameter tuning, KNN's performance remains relatively stable compared to other models. While its R^2 value improves slightly, it still lags behind XGBoost and Random Forest. This indicates that KNN's effectiveness for ROP prediction may be limited, especially when dealing with complex datasets with numerous variables.
  - *ANN:* The performance of ANN improves with hyperparameter tuning, but the improvement is less significant compared to XGBoost. ANN still struggles with achieving high R^2 values, suggesting that this model might require further tuning and possibly a larger dataset to improve its predictive accuracy for ROP.

- **General Observations:**
  - *Tuning Benefits:* Hyperparameter tuning generally improved model performance, particularly for XGBoost, underscoring the importance of parameter optimization in enhancing model accuracy.
  - *Well Variability:* The models' performance varied between the two wells, suggesting that different geological or operational characteristics might affect predictive accuracy. This highlights the need for tailored models or additional feature engineering to account for well-specific factors.

- *Model Robustness:* Random Forest emerged as the most robust model, consistently performing well with both simple split and cross-validation, making it a reliable choice for ROP prediction in varying conditions.

The evaluation results demonstrate the effectiveness of hyperparameter tuning and cross-validation in improving the performance of machine learning models for ROP prediction. While Random Forest consistently performs well across both scenarios, XGBoost shows significant potential with appropriate optimization, especially in complex datasets.

This suggests that for accurate ROP prediction, more advanced algorithms like XGBoost and Random Forest, capable of capturing complex interactions between features, may be preferable. However, it is important to note that the choice of the best model may depend on specific dataset characteristics and application requirements. Future research should focus on investigating the impact of various feature engineering techniques on model performance, exploring the use of deeper ANN architectures, and testing the models on a larger and more diverse dataset to assess their generalizability.

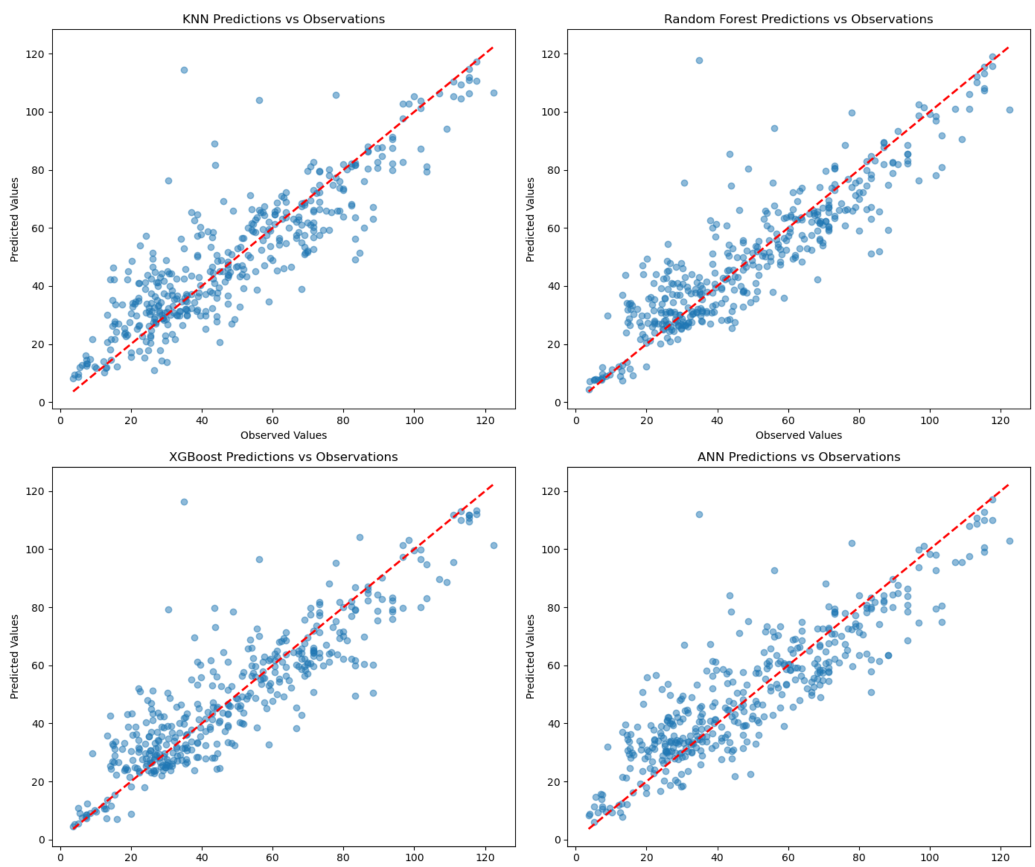


Figure IV.06: Visualization of the model's performances (Well 01).

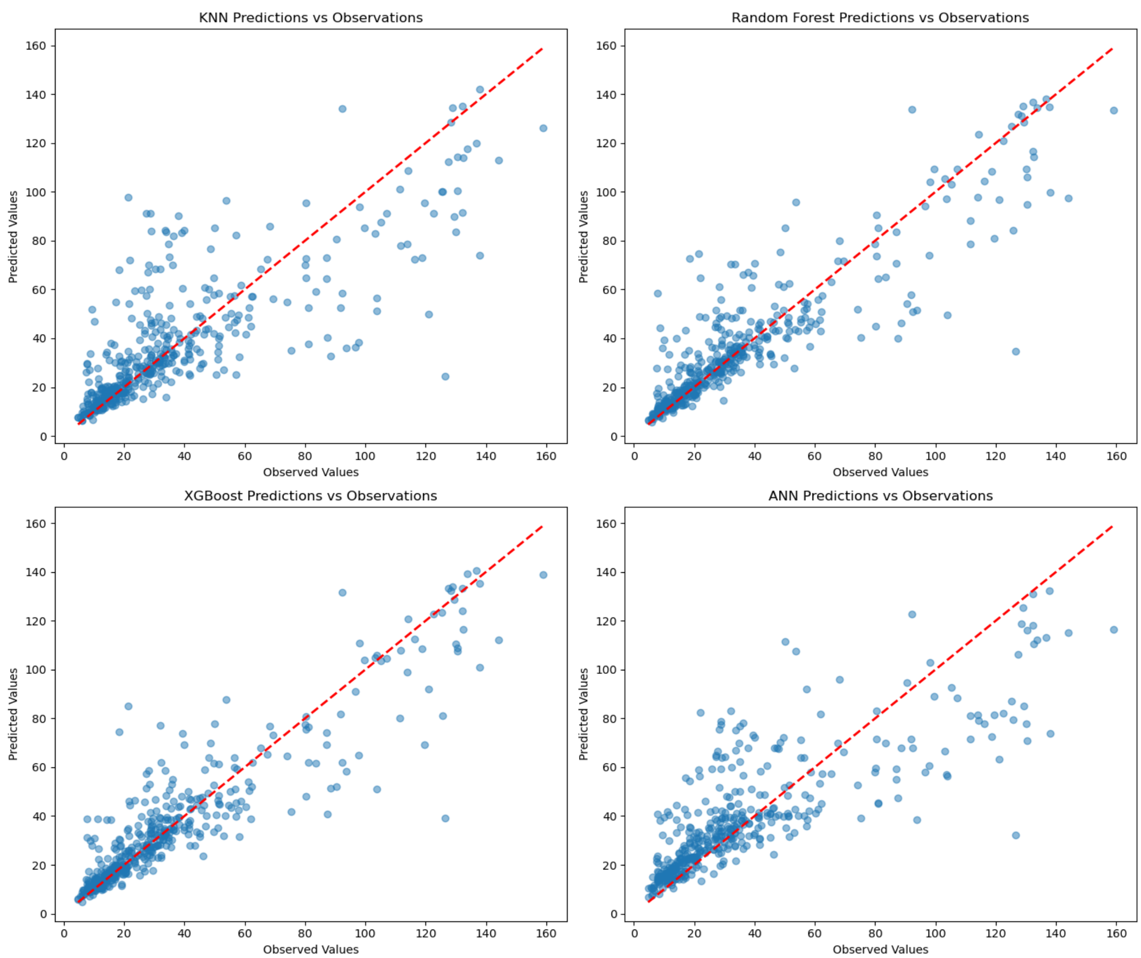
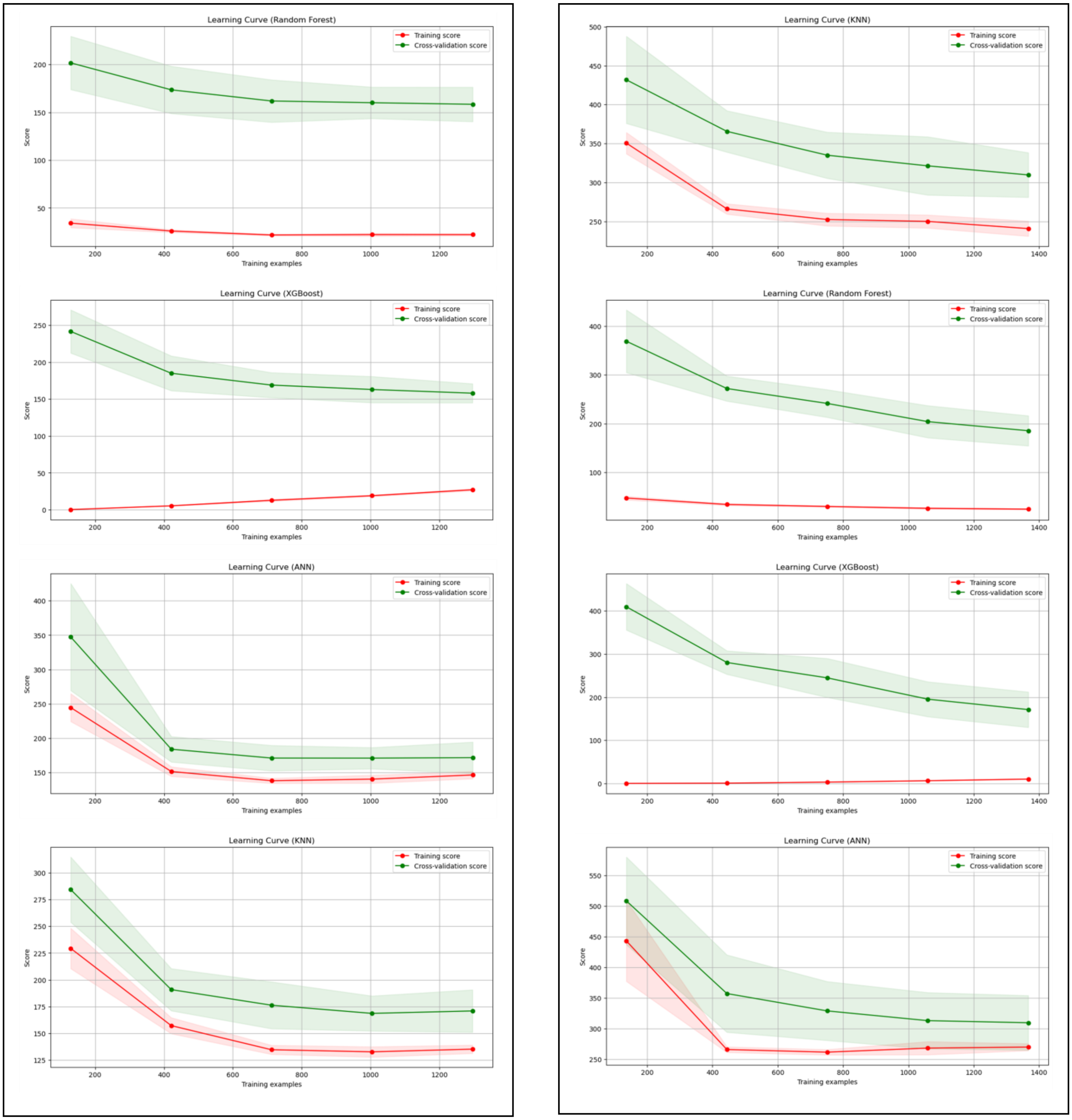


Figure IV.07: Visualization of the model's performances (Well 02).



(a) (b)

Figure IV.08: Learning curves of our models: (a) well1, (b) well 2.

Figure IV.08 shows the learning curves for four different machine learning models: K-Nearest Neighbors (KNN), Random Forest, XGBoost, and Artificial Neural Network (ANN). These learning curves are plotted for two different wells, labeled (a) and (b), to evaluate the models' performance in predicting the rate of penetration during drilling operations.

In both cases, the x-axis represents the training examples, and the y-axis represents the score or performance metric being used (likely some measure of error or accuracy). The learning curves for each model are plotted using two lines: one showing the training score (typically the error on the training data) and another showing the cross-validation score (the error on a held-out validation set).

For well (a), we can observe the following:

- The Random Forest model appears to perform the best overall, with the highest cross-validation score and the smallest gap between training and validation scores, indicating good generalization.
- The XGBoost model also performs reasonably well, with a slightly lower cross-validation score than Random Forest but still better than KNN and ANN.
- The KNN and ANN models have larger gaps between their training and validation scores, suggesting potential overfitting issues.

For well (b), the overall performance of the models is lower compared to well (a), indicating that this well may be more challenging to predict. However, the relative performance of the models is similar:

- Random Forest and XGBoost still perform better than KNN and ANN.
- KNN and ANN exhibit larger gaps between training and validation scores, indicating potential overfitting.

Based on these learning curves, the Random Forest and XGBoost models appear to be the most promising for predicting the rate of penetration in these two wells. However, it's important to note that these results may vary for different wells or datasets, and further evaluation and tuning of the models may be necessary for optimal performance.

6.2 Drilling Bit Selection Outcomes:

The bit selection algorithm utilized in this study employed four distinct machine learning models: K-Nearest Neighbors (KNN), Random Forest, XGBoost, and Artificial Neural Network (ANN). Each model was trained to predict the Rate of Penetration (ROP) based on historical drilling data. The performance of these models was evaluated using Root Mean Square Error (RMSE) and the coefficient of determination (R^2), both before and after hyper-parameter tuning and cross-validation.

- **K-Nearest Neighbors (KNN):**

The KNN model demonstrated moderate performance in predicting ROP. Before hyper-parameter tuning, the RMSE for Well 01 was 12.789, with an R^2 value of 76.352%. For Well 02, the RMSE was 18.730 with an R^2 of 65.576%. After tuning, slight improvements were observed in Well 01 (RMSE: 12.551, R^2: 77.222%), while Well 02 showed a marginal decrease in performance (RMSE: 19.220, R^2: 63.752%). This indicates that KNN can be sensitive to parameter settings and may require careful tuning for optimal performance.

- **Random Forest:**

The Random Forest model consistently outperformed the other models across both wells. Without tuning, the RMSE for Well 01 was 11.675 with an R^2 of 80.291%, and for Well 02, the RMSE was 13.973 with an R^2 of 80.839%. Post tuning, the performance remained stable, with a slight improvement in consistency (Well 01 RMSE: 11.697, R^2: 80.217%; Well 02 RMSE: 14.125, R^2: 80.422%). These results suggest that Random Forest is robust and effective in handling drilling data variability, making it a reliable choice for bit selection.

- **XGBoost:**

XGBoost also showed strong performance. Initially, the RMSE for Well 01 was 12.297 with an R^2 of 78.136%, and for Well 02, the RMSE was 14.111 with an R^2 of 80.460%. Hyper-parameter tuning further enhanced the results, reducing the RMSE to 12.056 and increasing the R^2 to 78.986% for Well 01. Well 02 saw significant improvement with an RMSE of 13.232 and an R^2 of 82.818%. XGBoost's ability to handle complex relationships in the data and provide high accuracy makes it a valuable tool for predicting ROP and selecting optimal drill bits.

● **Artificial Neural Network (ANN):**

The ANN model provided good results but was slightly less consistent compared to Random Forest and XGBoost. Before tuning, well 01 had an RMSE of 12.810 with an R^2 of 76.275%, and Well 02 had an RMSE of 18.785 with an R^2 of 65.373%. Post tuning, the model improved slightly, with Well 01 showing an RMSE of 12.593 and an R^2 of 77.073%, and Well 02 showing an RMSE of 18.697 and an R^2 of 65.697%. While ANN can capture non-linear patterns effectively, it requires careful tuning and ample data to achieve optimal results.

The application of these machine learning models for bit selection significantly impacts drilling performance. By accurately predicting ROP, the models enable more informed decisions regarding bit selection, leading to improved drilling efficiency and reduced operational costs. Specifically, Random Forest and XGBoost models have demonstrated high predictive accuracy, suggesting that their use can optimize bit selection, minimize drilling time, and enhance overall well productivity. The results indicate that integrating advanced algorithms into drilling operations can lead to substantial performance improvements, making the drilling process more efficient and cost-effective.

- **Bit Database:** Create a database or dictionary containing information about different bit types, their characteristics (e.g., diameter, tooth design, gauge, etc.), and their typical ROP performance under different conditions. This database can be based on manufacturer specifications, field data, or a combination of both.
- **Bit Selection Logic:** You'll define a logic or algorithm that uses the predicted ROP and the bit database to recommend the most suitable bit type for the upcoming drilling section. This logic could consider factors like:
    - **Target ROP:** If you have a desired ROP range for a specific formation, select a bit that is likely to achieve that range based on its historical performance.
    - **Formation Properties:** Consider the predicted ROP and the formation's hardness, abrasiveness, etc. Choose a bit that is known to perform well in that type of formation.

- **Drilling Conditions:** Factors like mud weight, rotary speed, and weight on bit influence bit performance. Choose a bit that's well-suited to the expected drilling conditions.

6.3 Discussion:

In the context of this project and based on the results presented, we proposed to use the Random Forest model for the following reasons:

●**Robust and stable performance:** The model results showed excellent and consistent performance in both wells, achieving the lowest (RMSE) and highest (R^2) coefficient of determination value compared to other models. This indicates that the random forest is able to provide accurate predictions of the rate of penetration (ROP).

●**Ability to handle diverse data:** The random forest model is able to handle a variety of inputs and complex factors, making it suitable for predicting drilling performance in different geologic conditions.

●**Stability and reliability:** The model showed high stability after adjusting parameters and using cross-validation, which means that the results are repeatable and reliable in different environments.

●**Ease of interpretation:** Although the random forest is a fairly complex model, it provides a good measure of the importance of features, which helps in understanding the key factors that affect the penetration rate and making informed decisions on drill head selection.

Based on these advantages, the random forest model is the optimal choice for improving the accuracy of penetration rate prediction and optimizing the drill head selection process, resulting in improved drilling efficiency and reduced operational costs.

In this study, we use a trained Random Forest model to predict the Rate of Penetration (ROP) value based on the input drilling data. This approach allows us to utilize the predictive capabilities of the model to make informed decisions on the selection of the optimal drilling bit. The detailed steps of the process are as follows:

●**Entering current drilling data:** Drilling data related to the current well is collected, which includes input variables such as drilling depth (TVD), rotation rate (RPM), torque (Torque), drill head weight (WOB), flow rate (Flow), and pump pressure (spp).

●**Predict the ROP value using the trained model:** This data is fed into the trained model, and the model calculates and predicts the expected ROP value for the current well.

●**Compare the predicted ROP value with historical data:** We compare the predicted ROP value with the performance of different drill bit in previous wells. We have a database that contains information on the performance of different drill bit in old wells, including the actual penetration rate achieved with each drill bit.

Table.IV.10: the performance of drill bits

Well Name	Drilling contractor	Phase 16" Interval	drill bit	ROP
RHAS QH-1 ST1	ENAFOR	479-2505(m)	TFF913S	42.34
RHAS-1	ENAFOR	470-2495(m)	Q609F	31.89
RHAQZH-2	ENAFOR	430-2229(m)	TFF913S	49.89
RHAQZH-1	ENAFOR	449-2290(m)	FX96R	34.32
RHAW-1	ENTP	603-2740(m)	MM96R	20.51

●**Analyze drill bit performance:** We analyze the historical data of different drill bit to see which drill bit has performed close to the predicted ROP value. The most appropriate drill bit is determined based on how close the actual ROP is to the ROP predicted by the model.

●**Selection of the optimal drilling bit:** Based on the previous analysis, the drilling bit that has shown better and more stable performance in similar geologic conditions, and has achieved a ROP value close to expectations, is selected. This ensures that the drill bit selected is best suited to achieve high drilling efficiency and minimize operational costs

●**Evaluate performance and continuous improvement:** The performance of the selected drill head is monitored and compared to expectations. If the actual results are in line with the predictions, the model can be considered effective. If there are significant deviations, the causes are analyzed and the model or data used is updated to improve the accuracy of future predictions.

The use of machine learning models can increase the accuracy of predictions compared to traditional methods, which contributes to improving the overall performance of drilling operations. In addition, selecting the optimal drilling head based on accurate predictions leads to reduced drilling time and operating costs, and continuous performance analysis can be utilized to help develop and improve prediction models, contributing to the advancement of technologies used in the drilling industry.

7 Conclusion:

This chapter evaluated the performance of four machine learning models for ROP prediction: KNN, Random Forest, XGBoost, and ANN. The results demonstrate that Random Forest consistently outperforms other models, both with and without hyperparameter tuning. However, XGBoost shows significant potential with appropriate optimization, particularly on complex datasets.

The findings highlight the importance of hyperparameter tuning and cross-validation for improving model performance. While Random Forest offers reliable predictions across different scenarios, XGBoost's adaptability with optimization suggests it may be a more powerful tool for complex drilling environments.

Future research should explore the impact of advanced feature engineering techniques, investigate the effectiveness of deeper ANN architectures, and evaluate model performance on a larger and more diverse dataset to assess generalizability. By leveraging the power of machine learning, we can develop more accurate and robust ROP prediction models, leading to enhanced drilling efficiency, cost optimization, and improved reservoir management.

In addition, drill bit selection is a key issue that can be optimized by using accurate predictions of the rate of penetration (ROP) from the considered models. Adopting the Random Forest model as the main ROP prediction model can provide stable and reliable performance, contributing to the selection of the most efficient drill bits for the specific geological conditions.

However, the potential of the XGBoost model should not be overlooked, especially in complex environments where parameter tuning can significantly enhance the accuracy of the predictions. By analyzing the historical performance of drill bits and comparing it to the predictions from these models, the drill bits selection process can be optimized to achieve the best penetration rates and reduce the time and costs associated with drilling operations.

By leveraging machine learning capabilities, more accurate and robust ROP prediction models can be developed, leading to improved drilling efficiency, better reservoir management, and significant operational cost savings. Achieving these goals will make the drilling industry more sophisticated and innovative, and enhance the ability to deal with complex geological challenges more effectively.

Conclusion

This research is aimed at optimizing the selection of drilling bit selection in drilling operations by reviewing the evolution of this process from traditional methods to the use of modern artificial intelligence techniques. Drill bits play a vital role in the success of drilling operations, as drilling performance and efficiency are highly dependent on the quality and suitability of the drill bit used. Traditional methods of selecting drill bits rely heavily on experience and empirical equations, which are limited by the availability of data and their ability to deal with different geological complexities.

The study addressed the latest applications of artificial intelligence (AI) techniques in the field of hydrocarbons, with a focus on rate of penetration (ROP) prediction using machine learning (ML) models. The performance of four machine learning models was evaluated: K-Nearest Neighbors (KNN), Random Forest, XGBoost, and Artificial Neural Networks (ANN). The results showed that the Random Forest model offers superior performance in terms of prediction accuracy and stability in different geological scenarios, while the XGBoost model showed promising potential when parameters are properly tuned, especially when dealing with complex data.

These results demonstrate the importance of parameter tuning and cross-validation in improving the performance of predictive models. The use of these techniques can enhance the accuracy of predictions and thus improve the selection of drilling bits, leading to improved efficiency of drilling operations and reduced operational costs. Considered machine learning models can provide accurate predictions that help in making informed decisions on the selection of the most suitable drilling bits, leading to improved overall drilling performance and significant financial savings.

Recommendation

Based on the findings of this study, we make the following recommendations:

- This study was done only on the technical layer of drilling and its results are effective only on the areas near the wells used in the study, so future studies may include the entire well or the scope of the study can be expanded to include a larger geographical area.

- For more accurate and generalizable future research, machine learning models should be created through collaboration between an expert in oil exploration and an expert in artificial intelligence.

- Creating a global model covering a large geographical area requires a huge amount of data that must be processed accurately in order to give more realistic and accurate results.

- Due to its superior performance in predicting the rate of penetration (ROP), Random Forest is recommended as the main model for analyzing drilling data and predicting penetration rates in various geological conditions.

- It is important to explore the potential of the XGBoost model deeper by tuning the parameters and using cross-validation techniques. This model can be a powerful tool in complex geological environments where it can provide accurate predictions after being optimized.

- Deeper and more complex artificial neural network (ANN) architectures should be studied and developed to improve the accuracy of predictions in changing geologic conditions. This can contribute to better results in complex geologic environments.

- Careful analyses should be conducted to assess the cost and benefit of applying AI technologies to drilling operations. This will help ensure an optimal return on investment and significant financial cost savings.

References

1. G. Gabolde, J.P. Nguyen. Drilling Data Handbook. 6th. Paris: French Institute For Petroleum Publish, 1999.

2. BORYCZLEO, P. Drilling bit selection and optimization in exploration well 6507/64A in the northland ridge aera. s.l.: University of Stavanger., 2012.

3. C.ALVAREZ, (s.d.), Drilling Bits and Hydraulics. s.l.: Quartz School for well site superviseur.

4. R. BOUDOUR. Drill bit. BOUMERDES: Drilling and production: IAP, 2013.

5. Forage, Projet de perfectionnement des superviseurs de. Outils, Parameters et Drill off test . s.l.: IAP, 2020, June.

6. Nguyen, Tan. Chapter 5 -Drill Bit. s.l.: Drilling Engineer. pp. 1-18.

7. BORYCZLEO, P. Drilling bit selection and optimization in exploration well 6507/64A in the northland ridge aera. s.l.: University of Stavanger., 2012.

8. Farr, Andrew. Trenchless Technology. [Online] Sandvik HDD, January 1, 2009. [Cited: May 27, 2024.] https://trenchlesstechnology.com/drill-bit-selection-simplifying-the-process./

9. Bit selection using field drilling data and mathematical investigation. M S Momeni, S Ridha, S J Hosseini, B Meyghani, S Emamian. s.l.: IOP Conference Series: Materials Science and Engineering, 2018, Vol. 328 012008.

10. Drill Bit Selection Optimization Based on Rate of Penetration: Application of Artificial Neural Networks and Genetic Algorithms. Ahmed K. Abbas, Amel H. Assi, Hayder Abbas, Haidar Almubarak, Mortadha Al Saba, s.l.: Society of Petroleum Engineers, 2019, Vols. 197241-MS.

11. HOUILI Mokhtar, HAMOUDI Mohammed. Etude de performance des outils de forage dans le drain horizontal de la phase 6. s.l. : Université KASDI Merbah Ouargla, 2014.

12. Okoli, U. Optimizing the Technical Limit Value of Drilling Operations with Proper Bit Selection and Hydraulics. s.l.: Society of Petroleum Engineers, 2013. pp. 2-3.

13. PH. GUTIERREZ, N. D. Beak-even method of investment analyses. s.l.: Farm and Ranch Series, 2009. p. Fact Sheet No 3.759.

14. Djamel Rezki, L. H. (2020). Rate of Penetration (ROP) Prediction in Oil Drilling Based on Ensemble Machine Learning. Springer Nature Switzerland, 537-549.

15. Erkam Guresen, G. K. (2011). Using artificial neural network models in stock market index prediction. Expert Systems with Applications, 10389-10397.

16. Karroubi1, E. B. (2021). Prediction of penetration rate in drilling operations: a comparative study of three neural network forecast methods. Petroleum Exploration and Production Technology, 11

17. Cheng1, H. L. Applications of Artificial Intelligence in Oil and Gas Development. s.l. : Arch Computat Methods Eng 28, 2021. pp. 937–949.

18. M. Shanker, M. H. (1996). Effect of data standardization on neural network training. Omega, 385-397.

19. "Artificial intelligence in oil and gas exploration and production: A review" (2020), Journal of Petroleum Science and Engineering, [190], 107019.

20. "The role of AI in seismic interpretation and reservoir characterization" (2021), SEG Technical Program Expanded Abstracts, [2021], 1-4.

21. "Machine learning in oil and gas exploration: A review" (2022), Petroleum Science and Technology, [40], 1653-1672.

22. "Predicting oil and gas production using machine learning: A case study" (2021), Journal of Petroleum Exploration and Production Technology, [11], 1173-1182.

23. "Data-driven reservoir management: An application of machine learning to optimize oil production" (2022), Energy [240], 122541.

24. "Using AI to improve well performance and reduce downtime" (2020), Oil and Gas Facilities, [3], 7-12.

25. "Artificial intelligence in well monitoring: A review of current applications and future opportunities" (2023), Journal of Natural Gas Science and Engineering, [109], 104429.

26. "Artificial intelligence in refining and petrochemicals: A review" (2021), Computers & Chemical Engineering, [143], 107125.

27. "Machine learning for process optimization in refineries: A case study" (2022), Chemical Engineering Research and Design, [179], 109-119.

28. "AI-powered inventory management in the oil and gas industry: A case study" (2020), Journal of Manufacturing Systems, [57], 24-32.

29. "Forecasting demand for petroleum products using machine learning: A review" (2023), Energy Economics, [105], 105918.

30. "The impact of artificial intelligence on oil and gas distribution and marketing" (2019), Energy Policy, [132], 485-495.

31. "Optimizing oil and gas supply chain using machine learning: A review" (2022), Journal of Cleaner Production, [337], 130496.

32. "AI-based logistics optimization for oil and gas transportation: A case study" (2021), Transportation Research Part C: Emerging Technologies, [127], 103085.

33. "The role of artificial intelligence in oil and gas logistics: A review" (2023), Journal of Petroleum Exploration and Production Technology, [13], 423-434.

34. "Artificial intelligence in oil and gas exploration and production: A review" (2020), Journal of Petroleum Science and Engineering, [190], 107019.

35. "The role of AI in seismic interpretation and reservoir characterization" (2021), SEG Technical Program Expanded Abstracts, [2021], 1-4.

36. "Machine learning in oil and gas exploration: A review" (2022), Petroleum Science and Technology, [40], 1653-1672.

37. "Predicting oil and gas production using machine learning: A case study" (2021), Journal of Petroleum Exploration and Production Technology, [11], 1173-1182.

38. "Data-driven reservoir management: An application of machine learning to optimize oil production" (2022), Energy [240], 122541.

39. "Using AI to improve well performance and reduce downtime" (2020), Oil and Gas Facilities, [3], 7-12.

40. "Artificial intelligence in well monitoring: A review of current applications and future opportunities" (2023), Journal of Natural Gas Science and Engineering, [109], 104429.

41. "Intelligent Well Planning and Optimization in Oil and Gas Industry" (2021), IEEE Access, [9], 27591-27600.

42. "Forecasting oil and gas production using machine learning: A review" (2023), Petroleum Science and Technology, [41], 1-25.

43. "Machine learning for production forecasting in the oil and gas industry: A review" (2021), Journal of Petroleum Exploration and Production Technology, [12], 755-769.

44. "Artificial intelligence in well monitoring: A review of current applications and future opportunities" (2023), Journal of Natural Gas Science and Engineering, [109], 104429.

45. "Artificial intelligence for well performance optimization and production enhancement in the oil and gas industry" (2021), Journal of Petroleum Science and Engineering, [199], 107917.

46 "AI-Based Predictive Maintenance for Oil and Gas Wells" (2021), IEEE Access, [9], 67313-67324.

47. "Artificial intelligence for automated drilling in the oil and gas industry" (2021), Journal of Petroleum Exploration and Production Technology, [12], 883-893.

48. "Autonomous drilling systems: A review of recent advances and future perspectives" (2022), Petroleum Science and Technology, [40], 1261-1279.

49. "Artificial intelligence in refining and petrochemicals: A review" (2021), Computers & Chemical Engineering, [143], 107125.

50. "Machine learning for process optimization in refineries: A case study" (2022), Chemical Engineering Research and Design, [179], 109-119.

51. "Artificial intelligence in the refinery process: A review" (2020), Journal of Petroleum Science and Engineering, [189], 107004.

52. "Predictive maintenance for oil and gas refineries using machine learning: A review" (2022), Journal of Petroleum Science and Engineering, [212], 109187.

53. "AI-powered predictive maintenance in the oil and gas industry: A case study" (2021), Journal of Petroleum Exploration and Production Technology, [12], 1193-1204.

54. "Quality control in the oil and gas industry using machine learning: A review" (2023), Petroleum Science and Technology, [41], 1-23.

55. "Artificial intelligence for quality control in the oil and gas industry: A review" (2021), Journal of Petroleum Science and Engineering, [200], 108024.

56. "AI-powered inventory management in the oil and gas industry: A case study" (2020), Journal of Manufacturing Systems, [57], 24-32.

57. "Machine learning for inventory management in the oil and gas industry: A review" (2021), Journal of Petroleum Exploration and Production Technology, [12], 1087-1099.

58. "The impact of artificial intelligence on oil and gas distribution and marketing" (2019), Energy Policy, [132], 485-495.

59. "Machine learning for demand forecasting in the oil and gas industry: A review" (2022), Journal of Petroleum Science and Engineering, [213], 109243.

60. "Price optimization in the oil and gas industry using machine learning: A review" (2023), Petroleum Science and Technology, [41], 1-20.

61. "Dynamic pricing in the oil and gas industry: A machine learning approach" (2021), Journal of Petroleum Exploration and Production Technology, [12], 979-988.

62. "AI-based logistics optimization for oil and gas transportation: A case study" (2021), Transportation Research Part C: Emerging Technologies, [127], 103085.

63. "Optimizing oil and gas supply chain using machine learning: A review" (2022), Journal of Cleaner Production, [337], 130496.

64. "The role of artificial intelligence in oil and gas customer relationship management" (2021), Journal of Petroleum Science and Engineering, [201], 108054.

65. "AI-powered customer service in the oil and gas industry: A review" (2023), Petroleum Science and Technology, [41], 1-19.

66 "Artificial intelligence for environmental monitoring in the oil and gas industry: A review" (2022), Journal of Petroleum Science and Engineering, [214], 109334.

67. "Machine learning for environmental monitoring and management in the oil and gas industry: A review" (2021), Journal of Petroleum Exploration and Production Technology, [12], 1215-1228.

68. "Artificial intelligence for safety and risk management in the oil and gas industry: A review" (2023), Petroleum Science and Technology, [41], 1-22.

69. "Machine learning for safety and risk assessment in the oil and gas industry: A review" (2021), Journal of Petroleum Exploration and Production Technology, [12], 1229-1242.

70. "The role of artificial intelligence in data analytics for the oil and gas industry" (2021), Journal of Petroleum Science and Engineering, [200], 108033.

71. "Big data analytics in the oil and gas industry: A review" (2022), Journal of Petroleum Science and Engineering, [212], 109168.

72. "Machine learning for cybersecurity in the oil and gas industry: A review" (2021), Journal of Petroleum Exploration and Production Technology, [12], 1243-1256.

Annex

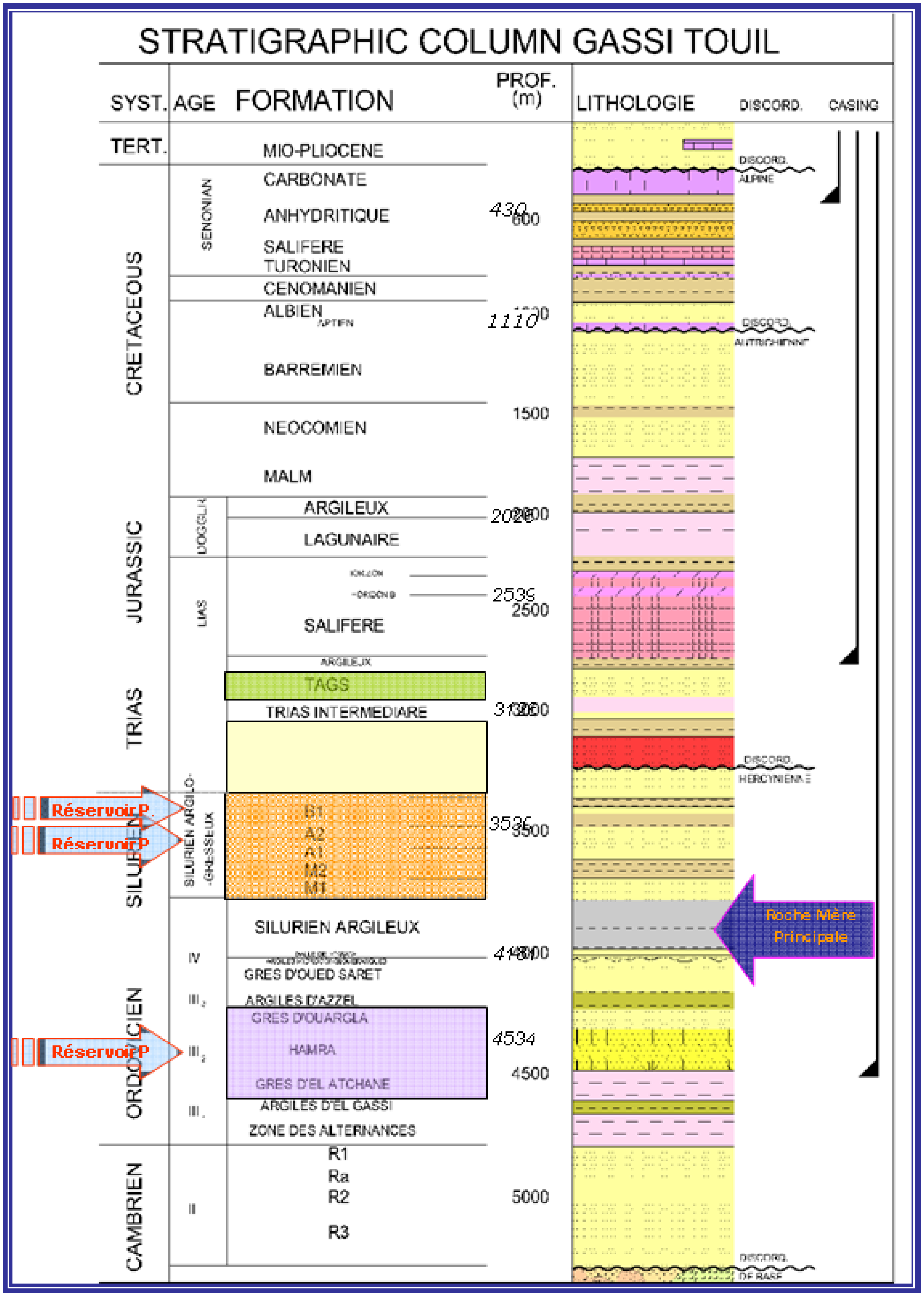


Figure.1 : Colonne litho-stratigraphique type de la région Rhourde Hamra- El Ledjmat.

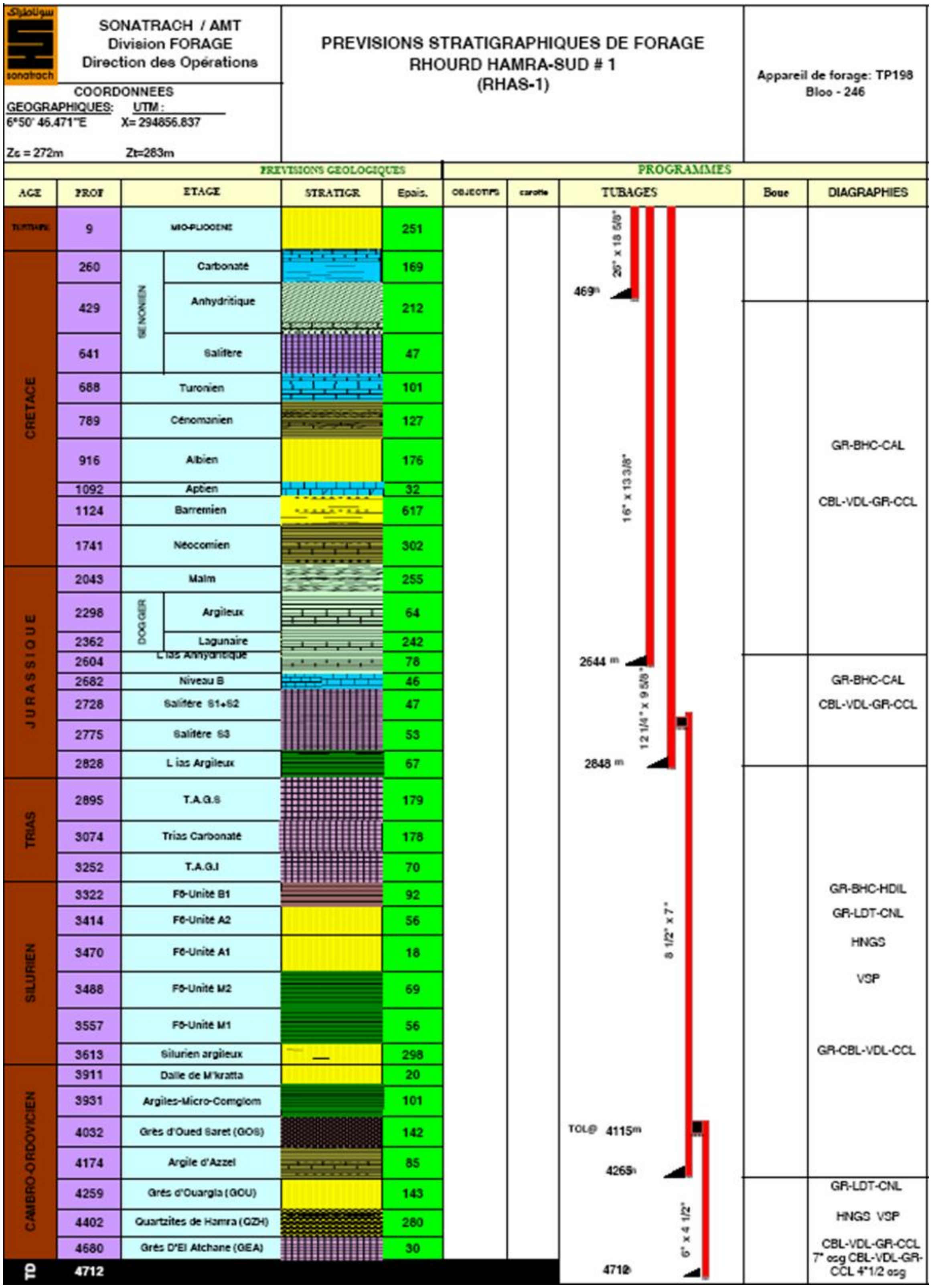


Figure.2: Drilling Program RHAS-1

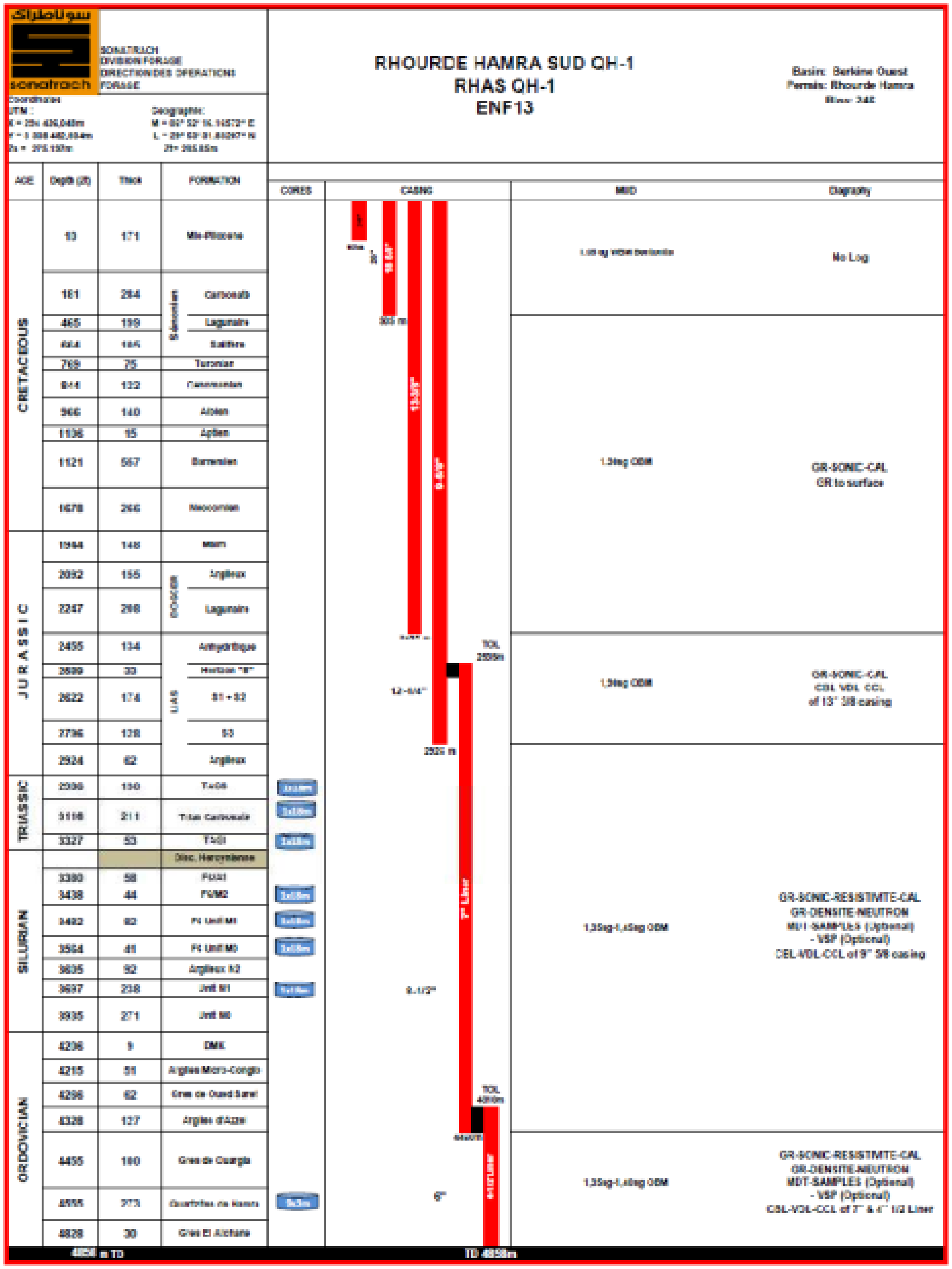


Figure.3: Drilling Program RHAS QH-1