REST Publisher; ISSN No: 2583-4800 (Online)

REST Journal on Advances in Mechanical Engineering

journal homepage: http://restpublisher.com/journals/jame/



Comparative Assessment of Machine Learning Models for Drilling Parameter Optimization in Offshore Wells

Sathiyaraj Chinnasamy*, M. Ramachandran, Muniyammal Sakthivel

REST Labs, Kaveripattinam, Krishnagiri, Tamil Nadu, India.

ARTICLE INFO

Article history:
Received: 01 October 2026
Accepted: 18 October 2026
Available online: 31 October 2026

Keywords:
Drilling Parameter
Penetration
Decision Tree Regression
XGBoost Regression

ABSTRACT

This research delves into optimising offshore drilling by leveraging sophisticated machine learning methodologies, with a particular emphasis on forecasting and scrutinising the Rate of Penetration (ROP) within the Volve Oil Field, North Sea. Since offshore drilling's inception in the late 19th century, it has cemented itself as a cornerstone of global energy supply, continually evolving through technological breakthroughs that bolster both its efficacy and safety. Addressing a pivotal challenge, this study concentrates on refining ROP prediction frameworks to streamline drilling activities, curtail operational expenditures, and mitigate ecological repercussions. The research's importance is underscored by its exhaustive exploration of drilling variables and their intricate interplay with ROP, integrating real-time drilling data alongside Computed Petrophysical Output (CPO) log records. Through machine learning-driven methodologies, the study aspires to construct robust predictive models capable of augmenting drilling efficiency and refining strategic decision-making in offshore ventures. Ultimately, this work extends the existing repository of knowledge on drilling optimisation while illustrating the tangible benefits of machine learning applications within the oil and gas sector. The methodological framework encompassed an in-depth analysis of drilling data leveraging three distinct machine learning paradigms: Decision Tree Regression, XGBoost Regression, and Histogram Gradient Boosting Regression. The study meticulously explored interdependencies among multiple drilling variables, such as Weight on Bit (WOB) and Surface Rotation Per Minute (SURF_RPM), alongside geological attributes like porosity (PHIF) and shale volume (VSH). To gauge predictive efficacy, the models were assessed through R-squared values, Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). Empirical results underscored XGBoost Regression as the most consistent performer, yielding robust predictive accuracy across both training (R^2 = 0.889) and testing (R^2 = 0.777) datasets. Notably, the investigation uncovered substantial correlations among drilling parameters, including a pronounced inverse relationship between water saturation (SW) and shale volume (VSH). The study inferred that machine learning methodologies, particularly XGBoost, establish a dependable framework for forecasting the Rate of Penetration (ROP) in offshore drilling scenarios. These insights pave the way for real-time drilling adjustments, enhancing operational efficiency while aligning with environmentally responsible practices in offshore resource extraction.

1. INTRODUCTION

Offshore drilling has entrenched itself as a pivotal force in the global energy framework, facilitating the extraction of hydrocarbons from beneath the ocean's depths. With the relentless surge in energy demand, the methodologies employed to tap into undersea oil reservoirs have undergone profound transformation over time. These shifts are shaped by leaps in technology, fluctuating economic landscapes, and an escalating consciousness regarding environmental repercussions.

*Corresponding author email-id: sathiyarajrsri@gmail.com
DOI: https://doi.org/10.46632/jame/5/4/1
Copyright @ 2026 Published by REST Publisher

The process of hydrocarbon exploration and extraction in offshore settings has not only redefined the contours of energy production but has also introduced an array of formidable challenges, necessitating strategic solutions to uphold sustainability [1-2]. Tracing its origins to the waning years of the 19th century, offshore drilling emerged when oil was first unearthed in shallow waters. The earliest known offshore oil well, established in 1896 off the Californian coast, stood on a timber-based platform. However, the sector did not witness substantial traction until the mid-20th century. A watershed moment arrived in 1947 with the discovery of oil in the Gulf of Mexico, setting the stage for increasingly sophisticated drilling mechanisms. By the 1960s, cutting-edge drilling innovations and the deployment of mobile drilling units empowered operators to venture into deeper marine territories, vastly broadening the scope of oil and gas extraction. This era ushered in an upsurge in offshore drilling ventures, particularly in hydrocarbon-abundant zones such as the North Sea and the Persian Gulf [3-5]. Offshore drilling rigs manifest in multiple forms, each meticulously engineered to cater to distinct environmental conditions and water depths. Among the most prevalent designs are fixed platforms, compliant towers, semi-submersible rigs, and drillships. Fixed platforms, firmly secured to the seabed, prove optimal for shallow waters, whereas compliant towers, constructed to endure severe meteorological forces, extend their utility into deeper regions. Semi-submersible rigs, partially buoyant and tethered below, deliver stability amidst turbulent seas. Meanwhile, drillships, outfitted with sophisticated dynamic positioning systems, navigate ultra-deep waters with remarkable agility, facilitating unparalleled mobility. Each category of rig is purpose-built to counteract the formidable challenges of the marine domain, ensuring precision, safety, and operational efficiency in drilling activities [6]. The drilling process unfolds in a sequence of intricate phases, commencing with exploration, wherein seismic surveys meticulously chart potential hydrocarbon reservoirs. Upon pinpointing a promising site, a drilling rig is mobilised to penetrate the ocean floor. This procedure necessitates a drill bit affixed to an elongated drill string, which rotates with formidable force to carve through subterranean layers. Throughout this operation, drilling mud circulates continuously, serving to regulate temperature, evacuate debris, and stabilise well pressure. Upon reaching the predetermined depth, extensive evaluations ascertain the site's production viability. If deemed commercially feasible, production apparatus is deployed to extract hydrocarbons. Cutting-edge advancements, including horizontal drilling and hydraulic fracturing, have revolutionised offshore extraction methodologies, granting access to previously untapped reserves with unprecedented efficiency [7-8]. Despite its technological strides, offshore drilling remains riddled with formidable challenges, each carrying profound operational and environmental implications. The oceanic environment is inherently volatile, with tempestuous weather patterns, forceful currents, and immense deep-sea pressures introducing substantial risks to drilling operations. Furthermore, the remoteness of many offshore sites exacerbates logistical hurdles, demanding meticulous coordination in transportation and supply chain logistics. The spectre of catastrophic accidents—ranging from blowouts to large-scale oil spills—looms persistently, threatening marine biodiversity and coastal livelihoods. The 2010 Deepwater Horizon calamity stands as a grim testament to these dangers, underscoring the imperative for uncompromising safety protocols and rigorous regulatory oversight [9-10]. Environmental ramifications sit at the heart of offshore drilling debates. Concerns surrounding oil spills, ecological disruption, and pollution fuel persistent apprehensions among conservationists and coastal populations. Consequently, regulatory frameworks have been meticulously crafted to impose stringent environmental assessments and enforce comprehensive safety mandates. In response, industry leaders are integrating progressive methodologies to curb ecological degradation, employing cutting-edge spill containment strategies and reinforcing safeguards for marine habitats. The fundamental dilemma persists: reconciling the insatiable global energy demand with the imperative to safeguard fragile oceanic ecosystems, a balance that necessitates perpetual dialogue and multi-stakeholder collaboration [11-12]. As the global energy paradigm continues its evolution, offshore drilling stands on the precipice of profound transformation. The relentless march toward renewable energy, coupled with an intensifying commitment to sustainability, is redefining industry trajectories. While conventional oil and gas extraction will persist as a cornerstone of global energy consumption, there is an unmistakable momentum toward the seamless incorporation of renewable technologies—wind and solar—into offshore infrastructures. Concurrently, advancements in carbon capture and storage may emerge as pivotal mechanisms, curbing the environmental toll of fossil fuel extraction. Faced with shifting market forces and mounting regulatory scrutiny, the offshore drilling sector is poised to embrace a future that prioritises sustainability and ethical energy practices [13-14]. With a legacy punctuated by groundbreaking technological strides, offshore drilling has long been a bastion of energy innovation. However, its progress is inextricably linked to a spectrum of formidable challenges, particularly in the realms of

environmental stewardship and operational safety. As the industry grapples with these intricate dynamics, the road ahead will hinge upon striking a delicate equilibrium between meeting global energy demands and mitigating ecological risks.

2. MATERIALS

This research utilises the 'Drilling Log Dataset' hosted on Kaggle (Elbashir, 2023). Comprising real-time drilling metrics alongside Computed Petrophysical Output (CPO) log records, the dataset pertains to well number 15/9-F-15, situated within the Volve Oil Field in the North Sea. It encompasses various parameters, including depth, weight on bit (WOB), surface rotation speed (SURF_RPM), penetration rate (ROP AVG), porosity (PHIF), shale volume (VSH), and water saturation (SW) [15].

- **Depth (Depth of Drilling):** Depth signifies the vertical or measured extent at which drilling operations are conducted, serving as a pivotal determinant of drilling performance. It governs critical parameters such as subsurface pressure, temperature variations, and geological formation responses. As drilling extends deeper, complications escalate—wellbore instability intensifies, fluid pressures surge, and mechanical stresses amplify. Effective depth control underpins strategic decisions regarding bit selection, drilling fluid composition, and casing design to optimise operational efficiency. Greater depths often introduce diverse lithological sequences, necessitating dynamic, adaptive methodologies. Precise depth tracking is imperative to mitigate differential sticking, excessive torque, and related drilling inefficiencies. Moreover, depth dictates critical choices concerning mud weight, bit specifications, and procedural refinements to uphold drilling stability. Accurate depth measurement is indispensable for geosteering precision, formation assessment, and optimising the wellbore trajectory, reinforcing its role as a crucial variable in drilling engineering.
- **WOB (Weight on Bit):** Weight on Bit (WOB) denotes the axial force exerted on the drill bit to facilitate penetration, directly influencing the Rate of Penetration (ROP). Generally, an increase in WOB accelerates drilling speed, yet excessive application risks bit degradation, premature mechanical failure, or structural damage to the drill string. Maintaining an optimised WOB ensures an equilibrium between penetration efficiency and bit durability. The ideal WOB is contingent upon multiple factors, including formation hardness, bit architecture, and drilling fluid characteristics. In softer formations, lower WOB minimises unnecessary wear, whereas harder strata necessitate greater force to achieve effective material removal. Drilling engineers actively modulate WOB in real time, leveraging sensor data and surface diagnostics to refine performance, curtail non-productive intervals, and enhance overall wellbore integrity.
- **SURF_RPM (Surface Rotation Per Minute):** Surface Rotation Per Minute (SURF_RPM) quantifies the drill string's rotational velocity, dictating bit performance and overall drilling dynamics. Elevating RPM accelerates the cutting process, thereby boosting the Rate of Penetration (ROP), especially in less resistant formations. However, excessive rotational speed may induce premature wear on both the bit and associated drilling apparatus, escalating risks of mechanical failure and severe vibrations. Determining the ideal RPM necessitates consideration of formation hardness, bit specifications, and operational limitations. Regulated RPM mitigates excessive thermal buildup, prolonging bit longevity while sustaining drilling efficiency. Contemporary drilling technologies employ real-time analytics to fine-tune RPM dynamically, enhancing energy utilisation and curbing mechanical degradation. When synchronised with Weight on Bit (WOB), meticulous RPM modulation optimises drilling performance, reinforcing borehole integrity and preempting subsurface complications.
- **VSH (Volume of Shale):** Volume of Shale (VSH) denotes the fraction of shale present within a geological formation, exerting a profound impact on drilling operations by altering porosity, permeability, and fluid retention characteristics. Elevated VSH signifies clay-dominant formations, which pose challenges such as wellbore instability, swelling tendencies, and potential drilling fluid contamination. In contrast, formations with minimal shale content typically exhibit greater structural stability, mitigating drilling hazards. Precise VSH evaluation informs the selection of optimal drilling fluid formulations, averting wellbore collapse. Geologists and drilling engineers

leverage gamma-ray logging and various petrophysical assessments to ascertain shale volume, refining drilling methodologies for maximised penetration efficiency. A nuanced understanding of VSH is indispensable in reservoir analysis, influencing hydrocarbon extraction efficacy. Strategic shale volume management facilitates seamless drilling execution while augmenting the precision of formation evaluation.

- **ROP AVG (Rate of Penetration Average):** The Rate of Penetration (ROP) Average serves as a crucial benchmark in drilling operations, reflecting the speed at which a drill bit advances through subterranean formations. A heightened ROP correlates with enhanced efficiency, effectively curtailing both time and expenses. Multiple factors govern ROP dynamics, including Weight on Bit (WOB), Surface Rotation Per Minute, drill bit selection, and geological composition. Achieving an optimal ROP necessitates a delicate equilibrium between acceleration, bit endurance, and borehole integrity. Excessive penetration rates risk destabilising the borehole, fostering cuttings buildup and differential sticking. Conversely, sluggish ROP prolongs drilling duration, inflating operational costs. Engineers leverage real-time monitoring to dynamically adjust drilling parameters, ensuring steady progress. Innovations in automation and predictive analytics further refine ROP strategies, driving efficiency while safeguarding wellbore stability.
- **PHIF (Porosity):** Porosity (PHIF) quantifies the fraction of void spaces within a rock formation, directly impacting its fluid storage capacity—be it hydrocarbons or water. Formations with elevated porosity generally signify superior reservoir potential, accommodating greater fluid volumes. However, effective porosity must be evaluated alongside permeability to determine actual fluid movement capabilities. Several geological factors, including grain size, sorting, compaction, and cementation, influence porosity variations. In drilling operations, precise porosity assessment underpins reservoir characterisation and completion strategy formulation. Logging tools such as neutron and density logs furnish real-time porosity data, guiding drilling decisions with accuracy. Comprehending porosity aids in selecting optimal drilling fluids, mitigating formation damage, and enhancing hydrocarbon recovery. Reliable porosity evaluation is indispensable for effective reservoir analysis and production forecasting.
- **SW (Water Saturation):** Water Saturation (SW) delineates the proportion of a formation's pore volume occupied by water, relative to hydrocarbons. This parameter holds significant weight in reservoir engineering, shaping hydrocarbon extraction efficiency. Elevated SW often correlates with diminished hydrocarbon reserves, whereas lower saturation levels signal more promising productivity. Precise SW evaluation informs well placement, production planning, and reservoir optimisation. Techniques such as resistivity and neutron logging facilitate the differentiation between water and hydrocarbons, enabling more refined decision-making. Drilling operations factor in SW values to fine-tune mud weights and minimise formation impairment. Managing water saturation proves pivotal in enhanced oil recovery strategies, ensuring optimal hydrocarbon extraction while mitigating excessive water output. A well-executed SW analysis fortifies reservoir management and long-term production viability.

3. OPTIMIZATION TECHNIQUES

Decision Tree Regression

Decision Tree Regression stands as a potent yet remarkably intuitive machine learning approach designed to predict continuous numerical values. Rooted in the principles of decision trees, this technique systematically partitions data into smaller subsets by evaluating the values of input features. Beyond its predictive power, it excels in offering interpretable decision pathways alongside visually comprehensible structures. As a supervised learning methodology, its versatility spans diverse domains—finance, healthcare, and marketing—where it effectively captures both linear and non-linear patterns within datasets [16].

Fundamentally, Decision Tree Regression constructs a hierarchical model resembling a branching tree, where internal nodes signify distinct features (or attributes), branches delineate decision-making rules, and terminal leaves denote predicted numerical outcomes. The tree formation unfolds through a recursive process, iteratively splitting data subsets with the objective of reducing variance or minimising mean squared error (MSE) within each segment. This refinement continues

until predefined constraints—such as a maximum permissible depth or a minimal number of samples per node—signal termination [17].

Decision Tree Regression primarily aims to curtail prediction errors, striving for optimal accuracy. A widely adopted criterion for node splitting is the Mean Squared Error (MSE):

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

where N denotes the total number of observations, y_i represents the actual values, and $\hat{y}_i$ signifies the predicted outputs. The tree partitions data by selecting the feature and corresponding threshold that yield the minimum MSE across the resulting child nodes, ensuring each split is as precise as possible.

The construction of a Decision Tree Regression model unfolds through a sequence of methodical steps. Initially, the algorithm determines the most optimal feature for partitioning the dataset, relying on the evaluation of Mean Squared Error (MSE). Upon selecting a feature, the dataset is split into smaller segments, with the algorithm persistently applying this division in a recursive manner. This iterative process proceeds until predefined stopping conditions are satisfied. Once the decision tree is fully structured, predictions for novel data points are generated by navigating from the root node down to a terminal leaf, adhering to decision pathways dictated by the given input attributes [18].

Decision Tree Regression boasts multiple benefits, such as its capability to process both numerical and categorical variables, straightforward interpretability, and minimal necessity for data preprocessing. Furthermore, it effectively captures intricate relationships among features. Nevertheless, it is not without drawbacks. Decision trees have a propensity to overfit training data, which undermines their ability to generalise effectively to new, unseen instances. Additionally, they exhibit high sensitivity to minor fluctuations within the dataset, potentially yielding structurally divergent trees under slight variations in input data [19].

Decision Tree Regression finds applications across diverse fields. Within finance, it plays a crucial role in credit evaluation and risk estimation. In the healthcare sector, it aids in forecasting patient prognoses by analysing multiple clinical indicators. Meanwhile, marketing experts leverage it for customer classification and anticipating sales trends based on behavioural insights. Its adaptability renders it ideal for situations necessitating continuous variable predictions [20].

In contrast to alternative regression techniques like linear regression or support vector regression, Decision Tree Regression distinguishes itself through its non-parametric approach. Unlike models that presuppose a predefined data distribution, it accommodates intricate patterns without demanding extensive feature transformations. Nevertheless, in cases where variable interactions adhere to a linear framework, linear regression may demonstrate superior efficacy. Moreover, ensemble techniques such as Random Forests and Gradient Boosting refine decision tree-based models, counteracting overfitting while enhancing predictive precision [21].

XGBoost Regression

XGBoost, an abbreviation for Extreme Gradient Boosting, stands as a formidable machine learning framework extensively leveraged for both classification and regression challenges. Functioning as an optimised variant of gradient-boosted decision trees, it is engineered for remarkable computational efficiency and predictive prowess. Developed by Tianqi Chen, XGBoost has garnered widespread acclaim, primarily for its capacity to process vast datasets, its resilience against overfitting, and its exceptional accuracy in predictions. Owing to these strengths, it has solidified its place as a staple tool among data scientists, particularly in high-stakes arenas like Kaggle competitions [22].

Fundamentally, XGBoost constructs an ensemble of decision trees, with each successive tree aiming to mitigate the errors left unaddressed by its predecessors. The algorithm capitalises on a method called boosting, wherein trees are iteratively appended, honing in on misclassified or inaccurately predicted instances from previous rounds. By fusing the outputs of multiple weak learners—decision trees in this case—XGBoost formulates a highly robust predictive system. Among its

defining attributes are regularisation mechanisms, seamless handling of missing data, and parallelised execution, all of which collectively enhance its speed and efficacy [23.

The fundamental aim of XGBoost revolves around minimising a predefined loss function while simultaneously integrating a regularisation component to counteract overfitting. The overarching mathematical representation of this objective function is formulated as:

$$\mathcal{L} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

Where:

- $\mathcal{L}$ denotes the comprehensive objective function.
- l represents the loss function, which quantifies the degree to which the model's forecasts align with the true values.
- y_i corresponds to the actual observed value, whereas $\hat{y}_i$ signifies the predicted counterpart.
- K designates the total count of trees constituting the ensemble.
- $\Omega(f_k)$ encapsulates the regularisation term associated with the kth tree, conventionally defined as:

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$$

In this equation, T symbolises the number of terminal leaves within the tree, γ functions as a regularisation hyperparameter, and λ governs the model's overall complexity by penalising large leaf weights.

XGBoost follows a sequential learning process, beginning with the construction of a rudimentary decision tree. Subsequently, the model evaluates residuals—discrepancies between actual and predicted outcomes—and generates an additional tree to approximate these residuals. This cycle persists iteratively, with each successive tree striving to diminish cumulative error. The ensemble of trees is aggregated through a weighted summation, wherein each tree's prediction is scaled by a learning rate, ensuring a controlled influence on the final output. Through this iterative refinement, XGBoost progressively enhances its predictive accuracy [24].

A key strength of XGBoost lies in its computational efficiency, particularly when dealing with extensive datasets. Leveraging parallel processing capabilities, it adeptly manages sparse data while maintaining performance. Furthermore, integrated regularization mechanisms curtail overfitting, reinforcing its reliability across diverse applications. Nevertheless, XGBoost is not without drawbacks—its structural intricacy often renders it less interpretable than more straightforward models. Moreover, achieving peak performance necessitates meticulous hyperparameter tuning, a process that is both resource-intensive and demands domain expertise [25].

XGBoost finds applications across a multitude of fields, spanning finance for credit evaluation, healthcare for forecasting diseases, and marketing for segmenting customers. Its adaptability ensures its relevance in any context necessitating regression or classification, solidifying its role as an indispensable asset in predictive modelling and machine learning [26].

In contrast to conventional regression techniques such as linear regression, XGBoost exhibits markedly superior performance, particularly when dealing with intricate datasets characterised by non-linear dependencies. Unlike linear regression, which presupposes a straightforward relationship among variables, XGBoost adeptly captures complex interactions and patterns. Furthermore, when placed alongside other ensemble approaches like Random Forests, XGBoost generally achieves higher efficacy, owing to its boosting mechanism that prioritises error correction, thereby enhancing its ability to generalise to new data. Consequently, while traditional models may offer greater interpretability, XGBoost stands out as a formidable alternative for attaining exceptional predictive accuracy [27].

Hist Gradient Boosting Regression

Histogram-based Gradient Boosting Regression stands out as an advanced machine learning framework that synergises the advantages of decision trees with the potency of gradient boosting. It excels particularly in regression scenarios, where the objective is to forecast continuous values based on provided input attributes. This technique surpasses conventional gradient boosting approaches by employing histograms to expedite the learning phase while simultaneously refining predictive accuracy. Given the ever-increasing necessity for swift and precise modelling solutions, Histogram-based Gradient Boosting Regression has found extensive application across diverse domains, spanning finance, healthcare, and marketing [28].

Gradient boosting, a formidable ensemble methodology, constructs predictive models progressively, aggregating multiple weak learners—primarily decision trees—to forge a highly resilient model. The histogram-based approach fine-tunes this mechanism by segmenting continuous variables into discrete bins, effectively simplifying computational complexity. This binning strategy not only diminishes training duration but also optimises memory consumption, rendering it particularly adept at handling vast datasets. By structuring data into histograms, the technique facilitates rapid calculation of gradients and Hessians, both of which are indispensable for iterative model enhancement during training [29].

The fundamental structure of gradient boosting adheres to the formulation:

$$F_m(x) = F_{m-1}(x) + \eta h_m(x)$$

where:

- $F_m(x)$ represents the model's predictive output at iteration m.
- $F_{m-1}(x)$ corresponds to the forecast derived from the preceding iteration.
- η denotes the learning rate, regulating the contribution of new learners.
- $h_m(x)$ signifies the newly integrated weak learner—typically a decision tree—appended during the mth iteration.

To iteratively minimise the loss function, which in regression tasks frequently manifests as mean squared error (MSE), predictions undergo refinement by aligning them with the gradients of the loss concerning prior estimations.

Histogram-based Gradient Boosting Regression operates through a sequence of fundamental stages. Initially, continuous numerical features undergo discretisation into finite bins, simplifying histogram construction. Each bin encapsulates a specific value range, streamlining statistical computations. During model training, gradients and Hessians corresponding to each bin are derived, serving as the foundation for decision tree construction. These trees iteratively predict the residuals of preceding models, progressively refining predictions until a termination criterion is satisfied. The final predictive framework emerges as a weighted aggregation of all the constructed trees [30].

A defining strength of Histogram-based Gradient Boosting Regression lies in its computational efficiency, particularly when handling extensive datasets. By leveraging histograms, the method substantially curtails processing complexity, leading to accelerated training durations. Moreover, it exhibits strong resilience against overfitting, especially when reinforced with regularisation strategies. Nevertheless, the technique is not without drawbacks—hyperparameter sensitivity poses a significant challenge. The choice of bin count, learning rate, and tree depth critically influences performance. Additionally, while the approach excels with structured datasets, its efficacy diminishes when confronted with unstructured data formats [31].

Histogram-based Gradient Boosting Regression finds extensive utility across diverse fields. In finance, it plays a pivotal role in evaluating creditworthiness and assessing financial risks. Within healthcare, it facilitates the prediction of patient prognoses by leveraging clinical datasets. Meanwhile, marketing experts harness its power for segmenting consumer groups and forecasting sales trends. Its capacity to process vast datasets while delivering precise predictions cements its status as a preferred analytical tool in these sectors [32].

When juxtaposed with alternative regression techniques, Histogram-based Gradient Boosting Regression distinguishes itself through its adaptability and predictive precision. Conventional linear regression models often falter when confronted with intricate, non-linear data patterns, whereas tree-based algorithms such as Random Forest frequently succumb to overfitting. In contrast, Histogram-based Gradient Boosting achieves an optimal trade-off between computational complexity and predictive strength, rendering it more adept at addressing a broader spectrum of analytical challenges. Serving as a formidable alternative to both linear and non-linear approaches, it not only enhances predictive accuracy but also boasts accelerated training efficiency [33].

4. ANALYSIS AND DISCUSSION

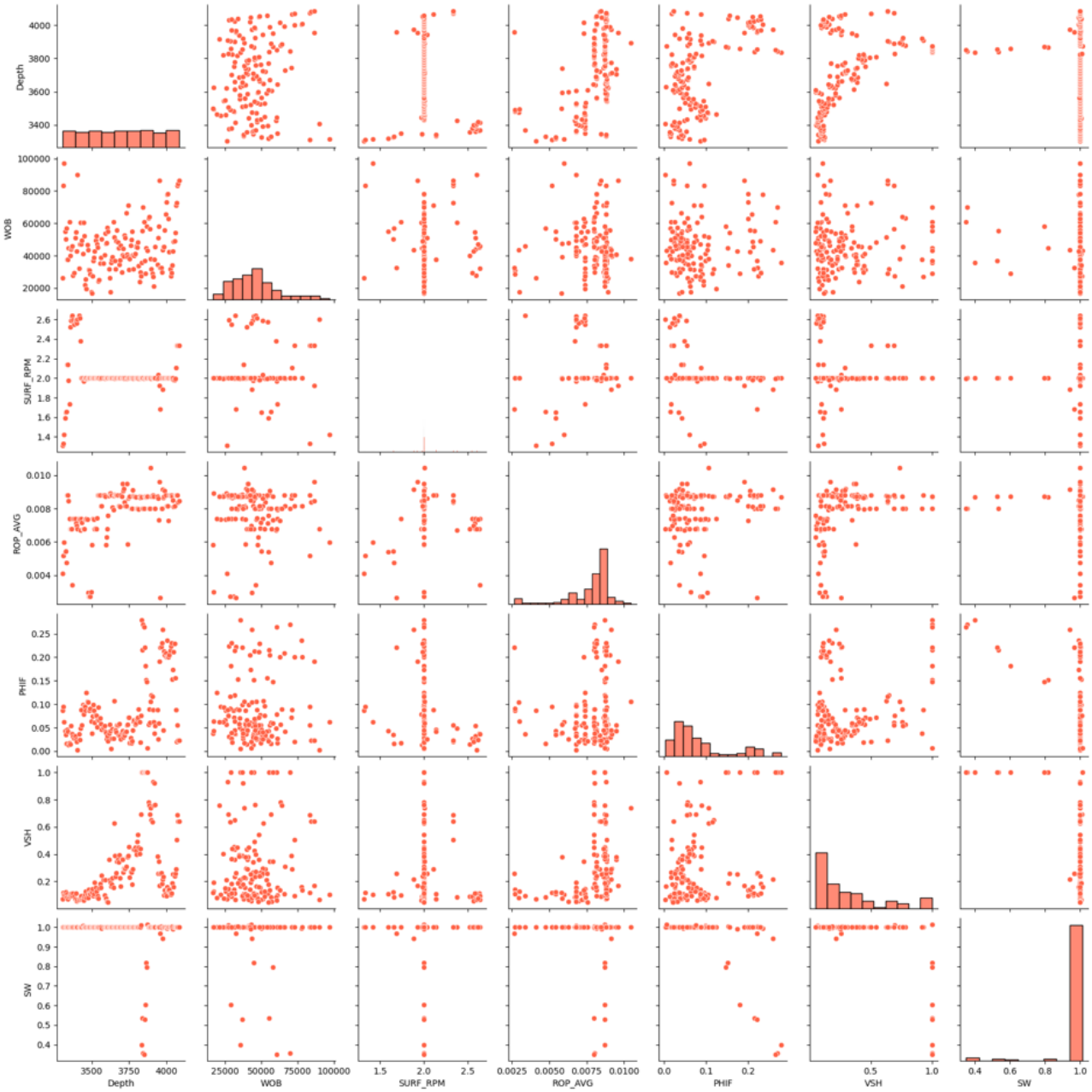


FIGURE 1. Effect of process parameters

Figure 1 unveils a multitude of intriguing trends and interdependencies within the drilling dataset. Examining depth-related

correlations, WOB (Weight on Bit) exhibits a largely stable pattern across varying depths, implying a consistent approach to drilling operations. Meanwhile, the surface RPM (SURF_RPM) remains strikingly uniform around 2.0, reflecting a standardised rotational speed setting throughout. The Rate of Penetration (ROP_AVG) fluctuates somewhat with depth, generally oscillating between 0.004 and 0.010, though sporadic anomalies emerge. This suggests a predominantly regulated drilling velocity, with intermittent modifications likely necessitated by geological transitions. Porosity (PHIF) values scatter between 0.05 and 0.25, subtly declining with increasing depth—a logical outcome as compaction intensifies in deeper strata. Water saturation (SW) presents a particularly compelling distribution, with most values clustering tightly near 1.0 (full saturation), yet exhibiting occasional abrupt plunges to approximately 0.3. Such drastic shifts may signify transitions between distinct lithologies or potential hydrocarbon-bearing intervals. The Volume of Shale (VSH) demonstrates a dispersed pattern across depths, reinforcing the notion of fluctuating lithological compositions throughout the wellbore. Diagonal histograms reveal that several variables deviate from normal distributions—WOB leans towards a right-skewed pattern, whereas SURF_RPM clusters sharply around discrete values, reinforcing the impression of controlled drilling settings. Interestingly, ROP_AVG's histogram hints at a slightly bimodal nature, potentially indicative of drilling through two primary formation types with contrasting characteristics. Cross-plots between variables expose further interrelations; for instance, a subtle inverse correlation surfaces between VSH and PHIF, aligning with geological expectations, as heightened shale content typically coincides with diminished porosity. Furthermore, WOB and ROP_AVG exhibit a mild positive correlation, implying that an increased weight on bit generally enhances penetration rates, albeit without a strictly linear relationship.

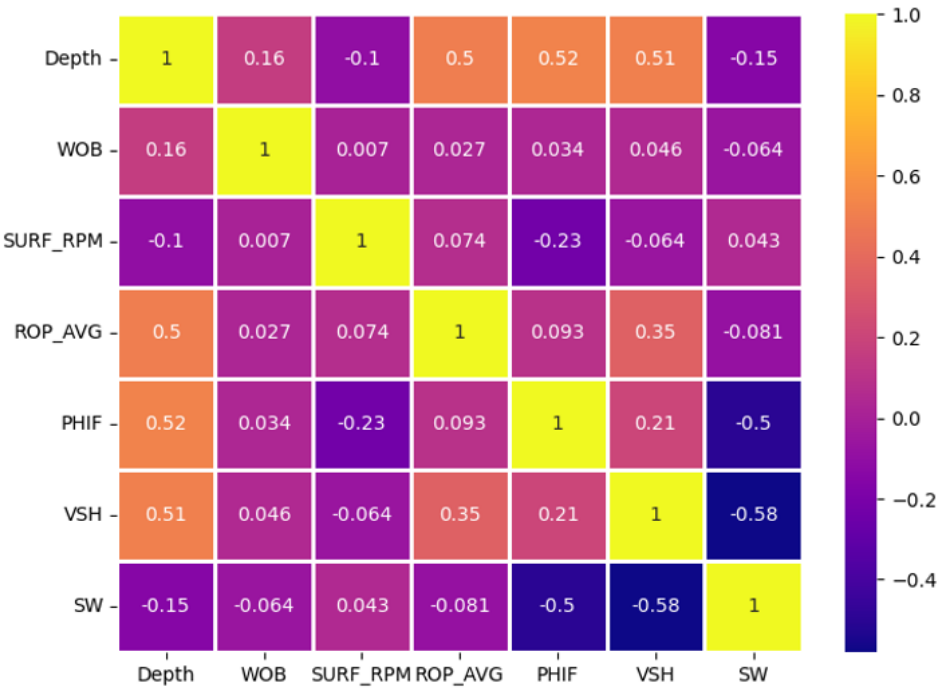


FIGURE 2. correlation matrix

The correlation matrix (figure 2) uncovers multiple noteworthy associations between drilling parameters and formation traits. Depth exhibits a moderately strong positive correlation with various factors, particularly PHIF (0.52) and VSH (0.51), implying that as drilling penetrates deeper, both porosity and shale volume tend to rise in tandem. This pattern likely signifies a structured evolution in formation properties with depth, potentially mirroring geological stratification or compaction effects. Weight on Bit (WOB) exhibits generally weak correlations across most parameters, with coefficients fluctuating between -0.064 and 0.16. This indicates that WOB remains largely detached from formation characteristics and is possibly governed more by operational considerations. The marginally positive correlation with depth (0.16) suggests only a faint inclination for increased weight at greater depths. Surface RPM maintains weak-to-moderate inverse

relationships with several parameters, most notably PHIF (-0.23). This negative trend may suggest that in formations with higher porosity, rotational speeds are deliberately reduced, perhaps to enhance borehole stability or optimise drilling efficiency. Its correlations with other parameters are minimal, underscoring the relative autonomy of RPM settings from formation characteristics. The dataset's most pronounced negative correlation emerges between SW (Water Saturation) and VSH (Volume of Shale) at -0.58, signifying that formations richer in shale tend to exhibit lower water saturation. This could be indicative of hydrocarbon presence or merely reflect the inherent inverse relationship between shale content and water retention within geological formations. Likewise, SW correlates negatively with PHIF (-0.5), suggesting that zones exhibiting greater porosity are often associated with reduced water saturation levels. Rate of Penetration (ROP_AVG) demonstrates moderate positive correlations with Depth (0.5) and VSH (0.35), implying that drilling velocity is somewhat responsive to these parameters. This trend may stem from either deliberate operational adjustments or the intrinsic reaction of the drilling assembly to varying formation conditions.

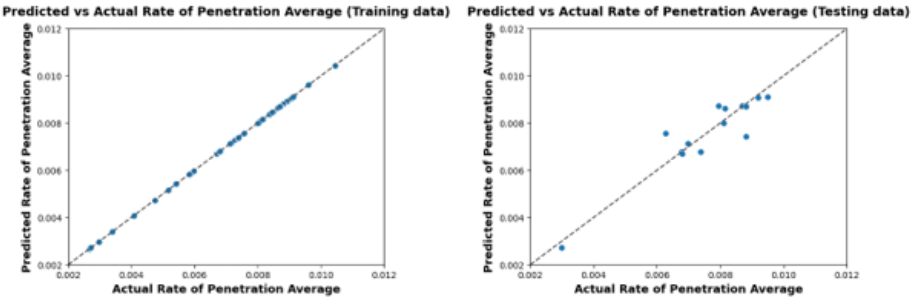


FIGURE 3. Decision Tree Regression (training and testing)

From image 3, the training data visualization illustrates a remarkable alignment between forecasted and actual ROP values, with nearly all points clustering tightly around the diagonal dashed line, which signifies an ideal 1:1 correspondence. This distribution, spanning approximately 0.002 to 0.012, forms a well-defined linear pattern, reinforcing the notion that the Decision Tree model has effectively internalised the underlying data structure with minimal divergence from true values. The even spread of points across the range further underscores the model's consistent predictive performance across various ROP magnitudes. Conversely, the testing data plot exhibits slightly altered behaviour. Although a clear positive relationship persists between predicted and actual values, there is noticeably greater dispersion around the ideal prediction line compared to the training data. The deviation is particularly pronounced in the mid-range (around 0.006-0.008), where the model occasionally overshoots or undershoots actual ROP figures. Such an increase in scatter is expected, as it reflects the model's adaptability to previously unobserved data. The disparity between training and testing performance signals a degree of overfitting—evident in the flawless fit seen in training data versus the broader spread observed in testing. This tendency is characteristic of Decision Trees, which excel at capturing intricate patterns in the training set yet may struggle to generalise as seamlessly to unfamiliar scenarios. Nevertheless, the overall trend in the testing data remains reasonably aligned with the diagonal, suggesting the model retains meaningful predictive utility. From a drilling optimisation standpoint, the model demonstrates substantial reliability in forecasting ROP, though practitioners should remain cognisant of heightened uncertainty in predictions when encountering novel geological formations or drilling environments.

The XGBoost regression plots (figure 4) illustrate the correlation between predicted and actual Rate of Penetration (ROP) averages for both training and testing datasets. The training plot exhibits strong predictive capability, yet a degree of dispersion is evident around the optimal prediction line, depicted by the diagonal dashed line. Data points predominantly align with this trajectory, reflecting an increasing trend, with ROP values spanning approximately 0.002 to 0.010. However, deviations emerge—lower ROP values appear slightly overestimated, whereas the mid-range displays a more dispersed pattern. This implies that while the XGBoost model effectively discerns fundamental trends, it retains a level of generalization rather than rigidly fitting the training data. The testing dataset yields a strikingly consistent outcome relative to the training set. The scatter distribution remains comparable, with data points closely adhering to the diagonal, particularly within the mid-range of ROP values (0.006–0.008). The reduced scatter in the lower ROP range suggests

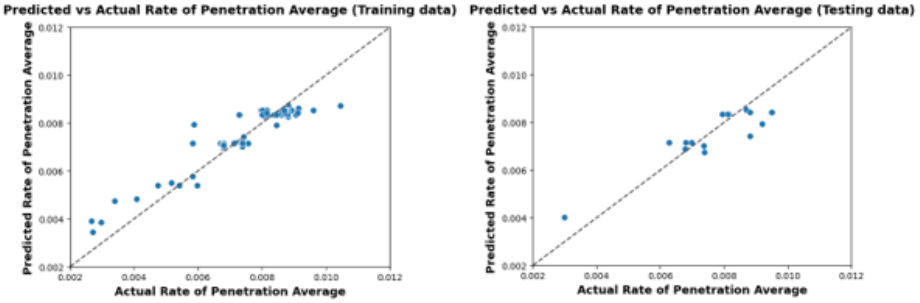


FIGURE 4. XGBoost Regression (training and testing)

the model has successfully captured underlying relationships, enhancing its applicability to unseen data. Notably, when juxtaposed with the Decision Tree model's prior performance, XGBoost demonstrates a more equitable balance between training and testing. Though the Decision Tree model may exhibit near-perfect alignment in training, XGBoost's steadier behaviour across datasets signals superior generalization, rendering it a more dependable option for real-world drilling optimisation. Its resistance to overfitting enhances its credibility, offering robust and reliable predictions across varied drilling conditions.

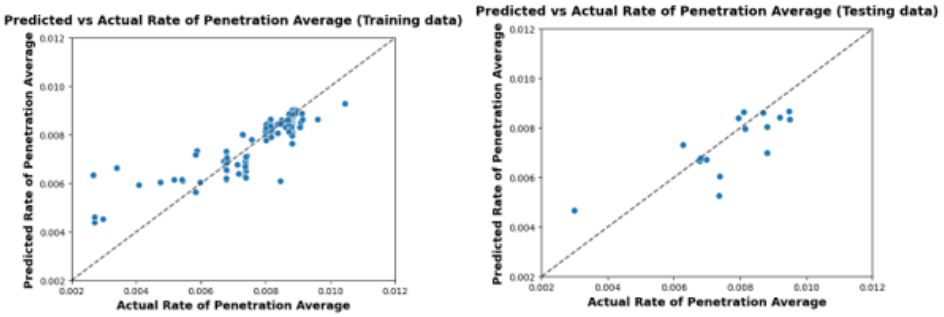


FIGURE 5. Hist Gradient Boosting Regression (training and testing)

Figure 5 presents Histogram Gradient Boosting Regression plots, juxtaposing the predicted and actual Rate of Penetration (ROP) Average for both training and testing datasets. In the training dataset plot, a moderate positive correlation is evident between predicted and actual ROP values, yet there exists a noticeable scatter around the ideal prediction trajectory, represented by the diagonal dashed line. The data distribution spans from approximately 0.002 to 0.010, with a pronounced clustering of points in the 0.007-0.009 interval. Notably, deviations from the ideal line emerge—lower ROP values are frequently overestimated by the model, whereas predictions in the mid-range exhibit substantial vertical dispersion. This dispersion signals that while the model has successfully discerned the overall pattern, it preserves a degree of flexibility rather than conforming rigidly to the training data. The testing dataset plot mirrors several traits observed in the training plot, albeit with a reduced number of data points, a natural consequence of the smaller test set size. The scatter maintains alignment with the diagonal trend line, with a predominant concentration of points within the 0.006-0.009 range. Some predictions diverge from the ideal trajectory—most notably, a lone point near the lower boundary (approximately 0.003) and several instances of spread among higher ROP values. The similarity between training and testing patterns suggests that the model exhibits sound generalization, avoiding extreme overfitting. Relative to previously assessed Decision Tree and XGBoost models, the Histogram Gradient Boosting approach introduces greater scatter in both training and testing predictions. Though this might superficially appear to indicate weaker performance, it may, in fact, reflect a more transparent representation of inherent uncertainties within ROP predictions. The model seemingly adopts a more conservative stance, steering clear of the excessive confidence that some alternative methods exhibit. This cautious predictive behaviour may prove advantageous in scenarios where grasping the uncertainty of predictions holds equal importance to sheer accuracy, though users should remain cognizant of its propensity to exhibit heightened variance in outputs.

The results presented in Table 1 demonstrate the comparative performance of the three regression models. The Decision

TABLE 1. Training and Testing Performance Comparison

	Train			Test		
Metric	**Decision Tree Regression**	**XGBoost Regression**	**Hist. Gradient Boosting Regression**	**Decision Tree Regression**	**XGBoost Regression**	**Hist. Gradient Boosting Regression**
Metric 1	1	0.888551	0.74448	0.898828	0.777355	0.546168
Metric 2	1	0.888551	0.74448	0.919514	0.790251	0.601229
Metric 3	0	2E-07	5E-07	2E-07	5E-07	1.1E-06
Metric 4	0	0.000487	0.000738	0.000498	0.000739	0.001055
Metric 5	0	0.000375	0.000467	0.000338	0.000605	0.000833
Metric 6	0	0.002076	0.00369	0.00142	0.001379	0.00215
Metric 7	0	2E-07	5E-07	2E-07	5E-07	1.1E-06
Metric 8	0	0.000293	0.000269	0.000221	0.000391	0.000785

Tree Regression exhibits an idealised fit on the training data, boasting R^2 and adjusted R^2 values of 1.0, alongside error metrics that vanish entirely—an unmistakable sign of overfitting. However, its performance on the test set, while still robust with R^2 at 0.899 and adjusted R^2 at 0.920, reveals minor yet non-negligible errors across MSE, RMSE, MAE, and MaxE. This divergence underscores the model's tendency to memorise rather than generalise, though its predictive capability on unseen data remains formidable. In contrast, XGBoost Regression strikes a more harmonious balance between learning and generalisation. Training results indicate R^2 and adjusted R^2 values of 0.889, with minimal error metrics (RMSE of 0.000487). On the test set, degradation is moderate, with R^2 declining to 0.777 and adjusted R^2 to 0.790, while error values (RMSE of 0.000739) rise slightly. This steadier transition between training and testing implies stronger generalisation. The Histogram Gradient Boosting Regression follows a markedly different trajectory, delivering the most restrained performance. With a comparatively lower R^2 of 0.744 on the training set, it undergoes the sharpest performance drop when evaluated on the test data, where R^2 and adjusted R^2 fall to 0.546 and 0.601, respectively. Correspondingly, its error metrics are higher, including a test RMSE of 0.001055 and MaxE of 0.002150. While these figures may seem less striking, they could indicate a more cautious and arguably realistic portrayal of predictive uncertainty. Evaluating the models collectively, a clear trade-off emerges between precision and generalisation. The Decision Tree prioritises accuracy but risks overfitting, XGBoost maintains equilibrium between predictive strength and adaptability, while Histogram Gradient Boosting adopts a conservative stance, potentially offering superior uncertainty estimation.

Examining the predictive capabilities of three regression models for Rate of Penetration (ROP) unveils critical insights. The Decision Tree model, despite achieving flawless accuracy in training ($R^2 = 1.0$), displayed clear overfitting tendencies. In contrast, XGBoost maintained a robust equilibrium between precision and generalisation, exhibiting consistent performance across training ($R^2 = 0.889$) and testing ($R^2 = 0.777$) datasets. The correlation heatmap underscored notable interdependencies among drilling parameters, particularly the link between depth and formation properties (VSH, PHIF). These findings highlight XGBoost as the most dependable model for practical ROP forecasting, striking an optimal balance between predictive accuracy and adaptability in real-world drilling scenarios.

5. CONCLUSION

Offshore drilling has become an indispensable pillar of global energy production, unlocking extensive hydrocarbon reserves hidden beneath the seabed. Originating in the late 19th century, this practice has undergone significant transformation, driven by technological innovation, heightened safety protocols, and a growing emphasis on environmental stewardship, cementing its role as a fundamental force in the contemporary energy landscape. This in-depth investigation into the interplay between drilling parameters and the Rate of Penetration (ROP) within offshore drilling operations has yielded profound revelations, facilitated by sophisticated machine learning methodologies. Scrutiny of the Volve Oil Field dataset uncovered intricate dependencies among key drilling variables, notably Weight on Bit (WOB), Surface Rotation Per Minute (SURF_RPM), and formation-specific attributes such as porosity (PHIF) and shale volume (VSH). The correlation assessment underscored significant associations, with a particularly pronounced inverse relationship observed between water saturation (SW) and shale volume (VSH), offering crucial insights for refining drilling strategies. The performance contrast among Decision Tree Regression, XGBoost Regression, and Histogram Gradient Boosting Regression uncovered notable distinctions. The Decision Tree model, despite attaining flawless accuracy on training data ($R^2 = 1.0$), exhibited overfitting when evaluated on unseen data. XGBoost, in contrast, demonstrated a well-balanced trade-off, preserving stability across both training ($R^2 = 0.889$) and testing ($R^2 = 0.777$) phases, indicating strong generalisation. Meanwhile, the Histogram

Gradient Boosting model, though more reserved in its predictions, provided meaningful insights into the uncertainty underlying its estimations. The implications of these findings are profound for real-world drilling practices. The established correlations between drilling parameters and the exceptional predictive capability of the XGBoost model offer a solid foundation for optimising ROP forecasts. Harnessing this insight enables enhanced drilling efficiency, cost reduction, and diminished environmental repercussions by refining parameter control. Furthermore, the study underscores the necessity of balancing precision with generalisation when choosing predictive models for drilling applications. Projecting ahead, this study establishes a groundwork for refining drilling optimisation methodologies. The demonstrated efficacy of machine learning in forecasting ROP hints at its applicability in broader drilling domains, including precise well trajectory design and dynamic drilling parameter modulation. Subsequent investigations might delve into synchronising these predictive models with live drilling data streams, paving the way for autonomous optimisation frameworks capable of responding to fluctuating formation properties and operational shifts. Furthermore, the approach formulated herein could be adapted for diverse offshore environments, enriching the overarching comprehension of drilling optimisation across varied geological landscapes.

REFERENCES

[1]. Al Husban, Moath. "Analysis of the Lebanese oil and gas exploration in the Mediterranean Sea: An overview and analysis of offshore platforms." International Journal of Advanced Engineering, Sciences and Applications 2, no. 1 (2021): 25-29.

[2]. Gubbala, Somasekhar. “TEMPO-Align: A Temporal Context-Aware Multi-Alignment Framework for Explainable Conformance Checking in Data-Rich Business Processes.” International Journal of Computer Engineering and Technology (IJCET) 15, no. 5 (2024): 1197–1214. IAEME Publication.

[3]. Saha, Sankhajit, Vikram Vishal, Bankim Mahanta, and Sarada Prasad Pradhan. "Geomechanical model construction to resolve field stress profile and reservoir rock properties of Jurassic Hugin Formation, Volve field, North Sea." Geomechanics and Geophysics for Geo-Energy and Geo-Resources 8, no. 2 (2022): 68.

[4]. Gaurav Saxena, "Real-Time Predictive Optimization Framework for Multi-Platform Digital Advertising Using Machine Learning and Multi-Criteria Decision Analysis," Journal of Software Quality Engineering and DevOps 1, no. 2 (2026): 1–14.

[5]. Kassem, Ahmed A., Souvik Sen, Ahmed E. Radwan, Wael K. Abdelghany, and Mohamed Abioui. "Effect of depletion and fluid injection in the Mesozoic and Paleozoic sandstone reservoirs of the October Oil Field, Central Gulf of Suez Basin: Implications on drilling, production and reservoir stability." Natural Resources Research 30, no. 3 (2021): 2587-2606.

[6]. Szymczak, Pat Davis. "Extended-reach drilling hits mainstream to squeeze difficult reservoirs." Journal of Petroleum Technology 73, no. 08 (2021): 35-37.

[7]. Li, Zhong, Li-yuan Xing, Jin-long Wang, Ren-jun Xie, Meng-bo Li, Zhi-ming Yin, Fu-xin Yu, Guo-ming Chen, and Xiu-quan Liu. "Temperature and pressure characteristics of the deepwater riserless drilling in the polar cold sea." Geoenergy Science and Engineering (2025): 213777.

[8]. Bello, Abdulwahab Muhammad, Stuart J. Jones, Jon Gluyas, and Khalid Al-Ramadan. "Impact of grain-coating clays on porosity preservation in paleocene turbidite channel sandstones: Nelson oil field, UK central North Sea." Minerals 12, no. 5 (2022): 555.

[9]. Ofei, Titus Ntow, Elie Ngouamba, Nils Opedal, Bjørnar Lund, and Arild Saasen. "Rheology assessment and barite sag in a typical North Sea oil-based drilling fluid at HPHT conditions." Korea-Australia Rheology Journal 35, no. 2 (2023): 81-94.

[10]. Lee, Hyungjoo, and Hunjoo P. Lee. "Formation lithology predictions based on measurement while drilling (MWD) using gradient boosting algorithms." Geoenergy Science and Engineering 227 (2023): 211917.

[11]. D’Almeida, Albino Lopes, Níssia Carvalho Rosa Bergiante, Geraldo de Souza Ferreira, Fabiana Rodrigues Leta, Cláudio Benevenuto de Campos Lima, and Gilson Brito Alves Lima. "Digital transformation: a review on artificial intelligence techniques in drilling and production applications." The International Journal of Advanced Manufacturing Technology 119, no. 9 (2022): 5553-5582.

[12]. Rathinavelu, G. (2025). AI Governance: Automatically Checking AI Systems Against the EU AI Act and NIST Rules. The Eastasouth Journal of Information System and Computer Science, 2(03), 428–438. https://doi.org/10.58812/esiscs.v2i03.1257

[13]. Gubbala, Somasekhar, and Praveen Kumar Kumbum. “Rewriting-Driven Reconfigurable Algebraic Petri Net Framework for Fault-Tolerant Asynchronous Federated Learning and Intelligent Vision Systems.” International Journal of Computer Science and Data Engineering 1, no. 3 (2024): 1–7. Sciforce Publications.

[14]. Parikh, Hirak, Mainavi Patel, Hardi Patel, and Gayatri Dave. "Assessing diatom distribution in Cambay Basin, Western Arabian Sea: impacts of oil spillage and chemical variables." Environmental Monitoring and Assessment 195, no. 8 (2023): 993.

[15]. Li, Qingchao, Qiang Li, Jingjuan Wu, Xianzhong Li, Hongbin Li, and Yuanfang Cheng. "Wellhead stability during development process of hydrate reservoir in the Northern South China Sea: Evolution and mechanism." Processes 13, no. 1 (2024): 40.

[16]. Mustaev, R. N., E. A. Lavrenova, V. Yu Kerimov, and R. A. Mamedov. "Peculiarities of Tertiary petroleum systems evolution under prograding shelf environment on the continental margin of the East Siberian Sea." Journal of Petroleum Exploration and Production Technology 11 (2021): 3617-3626.

[17]. Georgiev, Svetoslav V., Holly J. Stein, Judith L. Hannah, and Rolando di Primio. "Timing and origin of multiple petroleum

charges in the Solveig oil field, Norwegian North Sea: A rhenium-osmium isotopic study." AAPG Bulletin 105, no. 1 (2021): 109-134.

[18]. Vangala, Ramachandra Reddy, and Praveen Kumar Kumbum. "SPARC-Stream: A Sketch-Based Privacy-Aware Resilient Compression Framework for High-Throughput Graph Stream Analytics." International Journal of Cloud Computing and Supply Chain Management 1, no. 4 (2025): 1–8. Sciforce Publications.

[19]. Elbashir, Ahmed. 2023. "Drilling Log Dataset." Kaggle. Available at https://www.kaggle.com/datasets/ahmedelbashir99/drilling-log-dataset.

[20]. Karim, Rezaul, Md Khorshed Alam, and Md Rezaul Hossain. "Stock market analysis using linear regression and decision tree regression." In 2021 1st international conference on emerging smart technologies and applications (eSmarTA), pp. 1-6. IEEE, 2021.

[21]. Shi, Maolin, Weifei Hu, Muxi Li, Jian Zhang, Xueguan Song, and Wei Sun. "Ensemble regression based on polynomial regression-based decision tree and its application in the in-situ data of tunnel boring machine." Mechanical Systems and Signal Processing 188 (2023): 110022.

[22]. Balogun, Abdul-Lateef, and Abdulwaheed Tella. "Modelling and investigating the impacts of climatic variables on ozone concentration in Malaysia using correlation analysis with random forest, decision tree regression, linear regression, and support vector regression." Chemosphere 299 (2022): 134250.

[23]. Vangala, Ramachandra Reddy, and Praveen Kumar Kumbum. "HYPERION-Q: A Graph-Guided Self-Validating and Hardware-Aware Framework for High-Performance Query Processing." Journal of Data Science and Information Technology 1, no. 2 (2024): 1–7. Sciforce Publications.

[24]. Latif, Sarmad Dashti. "Developing a boosted decision tree regression prediction model as a sustainable tool for compressive strength of environmentally friendly concrete." Environmental Science and Pollution Research 28, no. 46 (2021): 65935-65944.

[25]. Gaurav Saxena, "Real-Time Predictive Optimization Framework for Multi-Platform Digital Advertising Using Machine Learning and Multi-Criteria Decision Analysis," Journal of Artificial Intelligence and Machine Learning 1, no. 4 (2023): 1–9.

[26]. Dandasi, Varun Venkatesh. 2025. "A Weighted Sum Method Approach to Ranking Machine Learning Techniques in Natural Language Processing." In Data Engineering Reimagined: Cloud, AI, ML & the Power of Analytics, vol. 1, 59–71. Sciforce Publications.

[27]. Gopinathan Rathinavelu (2024). A Scenario-Based Benchmark for Evaluating AI Safety Controls in Banking Operations: Metrics for Containment, False Intervention, and Recovery. International Journal of Future Innovative Science and Technology (IJFIST), Vol. 7 No. 4 (2024): International Journal of Future Innovative Science and Technology (IJFIST), pp. 13223-13232. https://doi.org/10.15662/IJFIST.2024.0704012

[28]. Shehadeh, Ali, Odey Alshboul, Rabia Emhamed Al Mamlook, and Ola Hamedat. "Machine learning models for predicting the residual value of heavy construction equipment: An evaluation of modified decision tree, LightGBM, and XGBoost regression." Automation in Construction 129 (2021): 103827.

[29]. Nitika, Rajesh Khanna, and Jaswinder Kaur. "Optimization of Modified T-Shape Microstrip Patch Antenna Using Differential Algorithm for X and Ku Band Applications." Microwave and Optical Technology Letters 60, no. 1 (2018): 219–229.

[30]. Patel, Richa Kamleshkumar, Eshant Aggarwal, Kamna Solanki, Omdev Dahiya, and Suman Avdhesh Yadav. "A Logistic Regression and Decision Tree Based Hybrid Approach to Predict Alzheimer's Disease." In 2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), pp. 722-726. IEEE, 2023.

[31]. Dong, Jianwei, Yumin Chen, Bingyu Yao, Xiao Zhang, and Nianfeng Zeng. "A neural network boosting regression model based on XGBoost." Applied Soft Computing 125 (2022): 109067.

[32]. Avanijaa, Jangaraj. "Prediction of house price using xgboost regression algorithm." Turkish Journal of Computer and Mathematics Education (TURCOMAT) 12, no. 2 (2021): 2151-2155.

[33]. Shahani, Niaz Muhammad, Xigui Zheng, Cancan Liu, Fawad Ul Hassan, and Peng Li. "Developing an XGBoost regression model for predicting young's modulus of intact sedimentary rocks for the stability of surface and subsurface structures." Frontiers in Earth Science 9 (2021): 761990.

[34]. Wang, Ruoran, Luping Wang, Jing Zhang, Min He, and Jianguo Xu. "XGBoost machine learning algorism performed better than regression models in predicting mortality of moderate-to-severe traumatic brain injury." World Neurosurgery 163 (2022): e617-e622.

[35]. Muniyammal Sakthivel, Arunambigai Ramesh, M. Ramachandran, Manjula Selvam, "Evaluation of Motor Vehicles Using the Weighted Product Model Method", REST Journal on Emerging trends in Modelling and Manufacturing, Vol 12(2) 2026, 1–13.

[36]. Shin, Hangsik. "XGBoost regression of the most significant photoplethysmogram features for assessing vascular aging." IEEE journal of biomedical and health informatics 26, no. 7 (2022): 3354-3361.

[37]. Abedi, Rahebeh, Romulus Costache, Hossein Shafizadeh-Moghadam, and Quoc Bao Pham. "Flash-flood susceptibility mapping based on XGBoost, random forest and boosted regression trees." Geocarto International 37, no. 19 (2022): 5479-5496.

[38]. Nitika, Jaswinder Kaur, and Rajesh Khanna. "Novel Monkey-Wrench-Shaped Microstrip Patch Sensor for Food Evaluation and Analysis." Journal of the Science of Food and Agriculture 102, no. 4 (2022): 1443–1456.

[39]. Tamim Kashifi, Mohammad, and Irfan Ahmad. "Efficient histogram-based gradient boosting approach for accident severity prediction with multisource data." Transportation research record 2676, no. 6 (2022): 236-258.

[40]. Qiu, Rangjian, Chunwei Liu, Ningbo Cui, Yang Gao, Longan Li, Zongjun Wu, Shouzheng Jiang, and Meng Hu. "Generalized Extreme Gradient Boosting model for predicting daily global solar radiation for locations without historical data." Energy

Conversion and Management 258 (2022): 115488.
[41]. Nagalla, Somesh. "Privacy-Preserving Emergency Data Mesh: A Homomorphic Encryption Approach to Multi-Agency Disaster Response Coordination." Journal of Computer Science and Technology Studies 7, no. 5 (2025): 778-785.
[42]. Sobolewski, Robert Adam, Médane Tchakorom, and Raphaël Couturier. "Gradient boosting-based approach for short-and medium-term wind turbine output power prediction." Renewable Energy 203 (2023): 142-160.
[43]. Tao, Hai, Salih Muhammad Awadh, Sinan Q. Salih, Shafik S. Shafik, and Zaher Mundher Yaseen. "Integration of extreme gradient boosting feature selection approach with machine learning models: application of weather relative humidity prediction." Neural Computing and Applications 34, no. 1 (2022): 515-533.
[44]. Sharma, J. K., V. D. Mishra, and R. Khanna. “Impact of Topography on Accuracy of Land Cover Spectral Change Vector Analysis Using AWiFS in Western Himalaya.” Journal of the Indian Society of Remote Sensing 41, no. 2 (2013): 223–235.
[45]. Nie, Peng, Michele Roccotelli, Maria Pia Fanti, Zhengfeng Ming, and Zhiwu Li. "Prediction of home energy consumption based on gradient boosting regression tree." Energy Reports 7 (2021): 1246-1255.
[46]. Li, Kaiqiao, Sijie Yao, Zhenyu Zhang, Biwei Cao, Christopher M. Wilson, Denise Kalos, Pei Fen Kuan, Ruoqing Zhu, and Xuefeng Wang. "Efficient gradient boosting for prognostic biomarker discovery." Bioinformatics 38, no. 6 (2022): 1631-1638.
[47]. Dandasi, V. V. "Application of ARAS Methodology in Supply Chain Performance Evaluation." International Journal of Cloud Computing and Supply Chain Management 1, no. 2 (2025): 1-7.
[48]. Nagalla, Somesh. "Technical Review Article: Self-Healing Lakehouse Manifests.", TIJER, 2025.
[49]. Vidhya Prasanth, Chitra Periyasamy, M. Ramachandran, Libiya Saravanan,”Biomedical Applications of Polymer-Composite Materials using Fuzzy ARAS Method” Journal on Materials and its Characterization, 4(4), 2025, 4-14.