Iraqi Journal of Oil & Gas Research, Vol. 03, No. 02 (2023)

Iraqi Journal of Oil & Gas Research

Journal homepage: https://ijogr.uotechnology.edu.iq

Penetration Rate Prediction Utilizing Machine Learning Technique for Rumaila Oilfield, Southern of Iraq.

[1]Ameen Kareem Salih, [1]Ali Khaleel Faraj, [2]Hassan Abdul Hadi Abdul Hussein and [3]Ali Nahi Abed Al-Hasnawi

[1]Oil and Gas Engineering Department, University of Technology-Iraq, Baghdad, Iraq
[2]Petroleum Department, College of Engineering, University of Baghdad, Iraq
[3]Drilling Branch, Petroleum Engineering, Islamic Azad University, Tehran - Iran

Article information

Article history:
Received: March, 07, 2023
Accepted: August, 28, 2023
Available online: October, 08, 2023

Keywords:
Machine learning;
Gradient Boosting;
Rumaila;
Artificial Intelligence

Corresponding Author:
Ameen Kareem Salih
150101@uotechnology.edu.iq.

Abstract

Cost estimation is an essential part of every drilling project's well planning, and it involves predicting the rate of penetration (ROP) accurately. The ROP represents the amount of time required to drill a given depth, and maximizing it helps minimize the costs associated with the drilling budget. However, predicting the ROP accurately is challenging because it depends on numerous variables, including drilling parameters, drilling fluid properties, and drilled formation characteristics.

One approach to improving the accuracy of ROP prediction is by using machine learning techniques, such as gradient boosting. In a recent study conducted in the Rumaila oilfield, the researchers used gradient boosting to predict the ROP based on drilling operation parameters and drilling fluid properties for two wells used for training and testing and one well used for implementation. The results of the study showed that gradient boosting was successful in predicting the ROP, with R2 training and testing values of 0.9947 and 0.8611, respectively. This means that the model was highly accurate and could be used to improve cost estimation in drilling projects.

Overall, the use of machine learning techniques such as gradient boosting can help enhance the accuracy of cost estimation in drilling projects by predicting the ROP more accurately, minimizing the costs associated with the drilling budget, and improving the overall efficiency of the drilling process.

DOI: http://doi.org/10.55699/ijogr.2023.0302.1042, Oil and Gas Engineering Department, University of Technology-Iraq

Introduction

Intricate drilling is the process of removing numerous types of rocks to reach the desired depth, the most crucial aspects in drilling that affects cost is penetration rate, often known as drilling speed. The length of time required to drill the well has the potential to significantly increase drilling expenses, [1]. Thus, one of the most important goals of drilling engineers is to reduce drilling time, [2-4]. The rate at which penetration occurs (ROP) is the main variable influencing drilling time, [5].

Rate of penetration (ROP) refers to the amount of rock or formation a drill bit can cut in a specific unit of time. Various factors such as mud properties, formation properties, depth, torque, WOB, RPM, Q, and SPP affect ROP. Optimal adjustments of these variables can enhance drilling efficiency and minimize expenses. It is essential to consider all these factors to ensure that drilling operations are successful and profitable, [6-10]. Some of these aspects, such as the formation features (porosity and lithology), are uncontrollable, while others, such as torque, weight on bit (WOB), rotation speed (RPM), and flow rate, are controllable. Most factor that affects the drilling rate is shown in figure 1.

There have been numerous attempts to model ROP using mathematical equations and statistical techniques, but they have been unsuccessful due to the great complexity of the ROP model or issues that can arise when results are generated in the lab or using insufficient field data, [3,11,12]. The goal of this work is to present research for ROP prediction in the southern Iraqi oil field of Rumaila using an intelligent model. ROP was determined by an intelligent model, which also demonstrated the consequences of different drilling parameters as well as its dependability and limitations.

Rumaila oil field is a supergiant oilfield that was discovered in 1953 contain two domes (south and north). The field is roughly located between latitudes 47° 14' and 47° 19' and longitudes 30° 13' and 30° 24'. Around 50 kilometers to the west of Basra city, on an area of 1800 km2 as shown in figure 2, [13,14]. The South of Rumaila field building was constructed in 1953 and placed into use in 1954. The Lower Cretaceous Zubair sandstone (Main Pay) was the intended focus. The North dome was then drilled in 1959, and it was discovered that the Main Pay and Mishrif carbonate, both from the middle Cretaceous, are the most prolific reservoirs, while the sandstone below was wet, [15].

1.1. Review

Intelligent models have gained more and more attention in recent years for its accuracy in calculating various drilling issues. In order to forecast ROP, Ahmed investigated the propensity of four widely utilized computational intelligence techniques (ANN, ELM, support vector regression (SVR), and least square support vector regression (LSSVR)), [17]. Shi in 2016 made use of upper-layer solution-aware techniques and extreme learning machines (ELM) to successfully forecast the ROP, [18]. Artificial neural networks (ANNs) were used by Azar in 2017 to build the ROP model, and the results showed how effective ANNs are as a tool for cutting costs, speeding up processes, and boosting structural reliability, [19].

Also, researchers are working on developing a ROP model utilizing the random forest technique, including the input characteristics RPM, WOB, flow rate, and uniaxial compressive strength (UCS), [20]. Bodaghi created the ROP model in 2015 using support vector regression that was enhanced by the Cuckoo search method and genetic algorithm, [21]. Hegde constructed a number of models based on the formation and contrasted them with data-driven models. Their conclusion demonstrates that data-driven models perform better, [22]. The application of several machine learning algorithms by the authors to forecast ROP while drilling in a specific formation has been informative, [23].

And though, in order to predict the rate of penetration (ROP) in drilling operations, artificial neural networks were developed and trained with four evolutionary techniques for accurate forecasting [24]. Regression issues have been widely addressed using the RBF neural network by authors, [25]. In 2020, Doaa developed a neural network model to forecast the rate of penetration for an Iraqi oil field. The model incorporated various parameters, including well depth, drilling fluid input, bit rotation speed, weight on bit, standpipe pressure, and bit size. The model's accuracy was evaluated based on its ability to estimate the rate of drilling penetration. The results showed that the neural network model provided excellent accuracy in predicting the rate of penetration. This forecasting tool could potentially benefit the oil and gas industry by optimizing drilling operations and reducing costs, [26]. As a result of early convergence, the PSO method is prone to being caught in local optimization, [27]. Due to its straightforward form and straightforward implementation, a nature-inspired optimization method has also attracted significant research attention, [28].

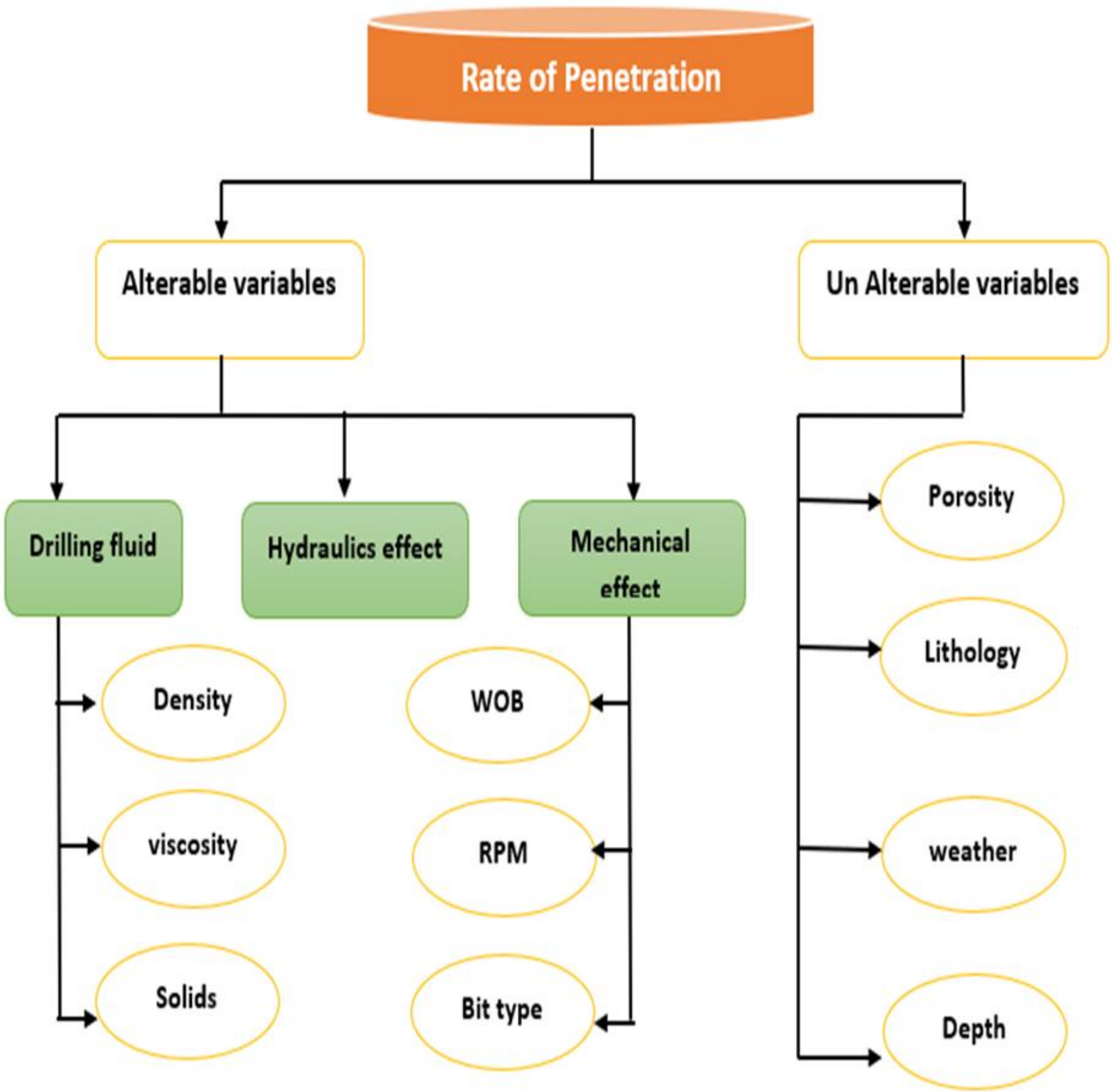

Figure 1: The most paramount factor that affects ROP.

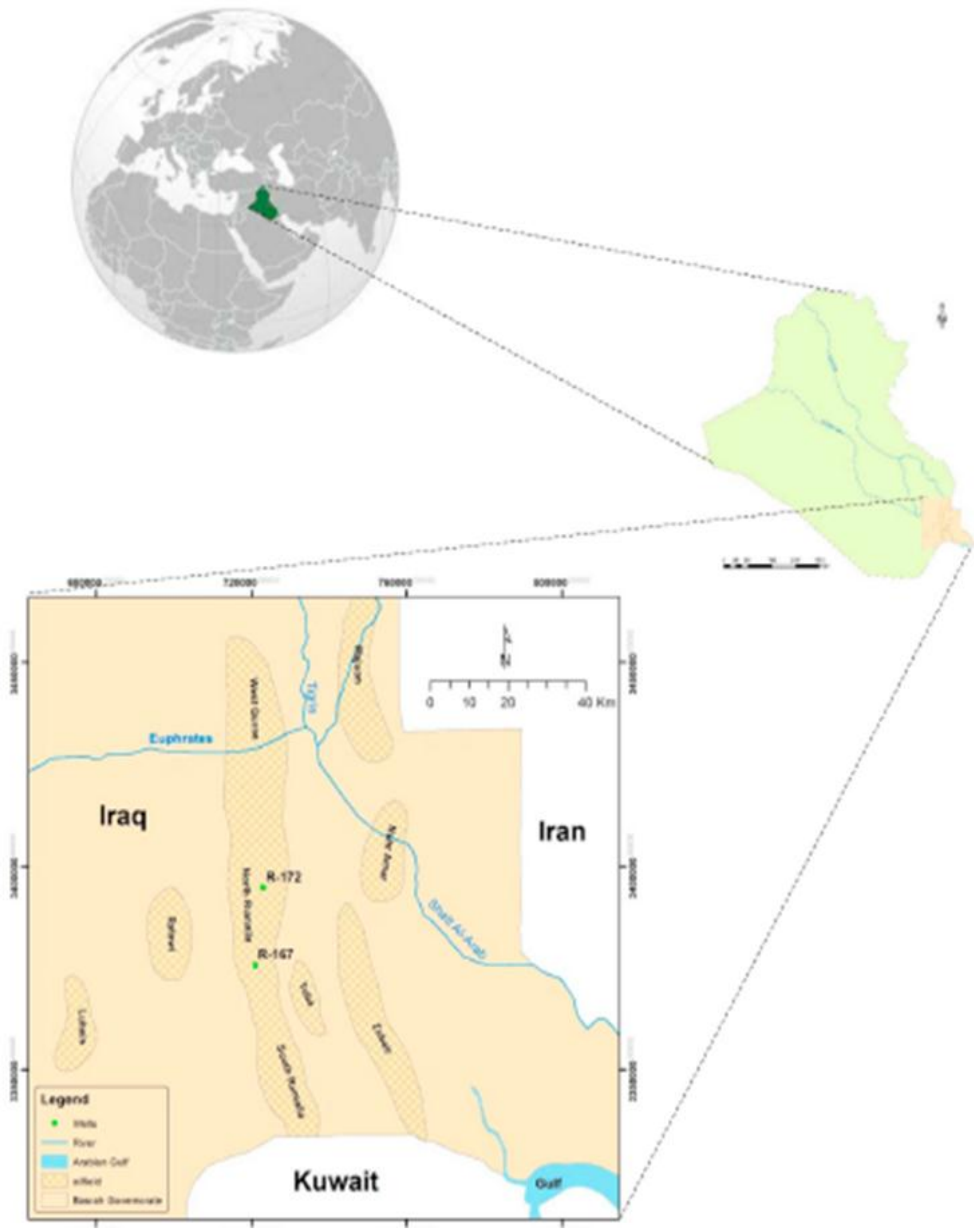

Figure 2: The location map provides a visual representation of the area of study, including geographical features, boundaries, and key landmarks [16].

1.2. Gradient Boosting

Gradient Boosting is an ensemble machine learning technique that combines multiple weak models to form a strong predictive model. It has gained popularity due to its ability to handle various data types and its customizability. The method involves iteratively adding new models to the ensemble and minimizing a loss function, such as mean squared error or binary cross-entropy, through gradient descent. Gradient Boosting can be used for classification, regression, and ranking problems, [29-31]. Gradient Boosting can be used for both Classification and Regression.

Basically, Gradient Boosting involves three elements, [32,33]:

1- A weak learner is a machine learning model that has modest predictive power, but is useful for building more complex models. Decision trees are a popular choice for weak learners in gradient boosting, although other models can also be used. The goal is to iteratively improve the ensemble by adding weaker learners [34].

2- A loss function is a crucial element in machine learning that determines how well the model fits the data. By minimizing the error, it helps in achieving better accuracy. Different types of loss functions are used for various purposes, such as regression and classification tasks.

3- A sequence of models, where each new model focuses on correcting the mistakes of the previous models. Gradient boosting is a popular form of boosting algorithm that uses gradient descent optimization to minimize the loss function. Boosting is often used in machine learning for tasks such as classification and regression.

When building predictive models, it's important to account for the possibility of incorrect predictions. By giving greater weight to incorrect predictions, we can train models that are better able to handle difficult cases and make more accurate predictions overall. One technique that can be used to accomplish this is gradient boosting. This involves training a sequence of models, each of which is designed to improve upon the predictions of the previous model. The process works by gradually reducing a loss function, which measures the error between the predicted values and the actual values. Overall, the goal of gradient boosting is to minimize the loss function, just as in an artificial neural network model where weights are tuned to minimize the error between predicted and actual values. By giving greater weight to difficult cases and focusing on improving the accuracy of our models over time, we can build better predictive models that are more effective at handling complex data sets and making accurate predictions, [32,35 ,36].

The predictions of many models are integrated in gradient boosting, as opposed to neural network models, where the goal is to minimize a loss function in a single model. Gradient boosting thereby makes use of some of the random forest/extra tree hyperparameters as well as other hyperparameters like learning rate, loss function, etc. that are utilized in an ANN model, [37-40].

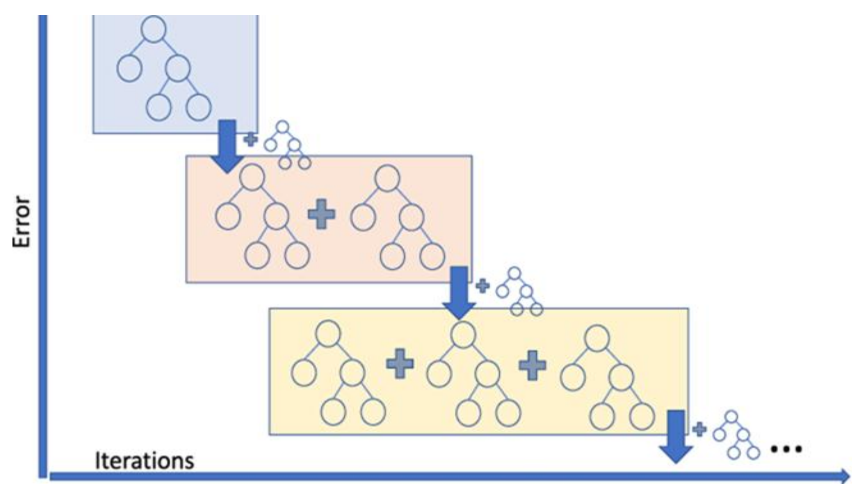

Figure 3: A schematic illustration of gradient boosting regression [39].

1. Methodology

Data gathering: data were collected of oil wells that were drilled in the Rumaila oil field, which included factors related to the drilling operation and the characteristics of the drilling fluid table 1. Figure (4) shows the relationships between the rate of penetration and the other parameters that were used in modeling.

Table 1: The import data using in this study.

index	count	mean	min	max
TVD(m)	4866	1257.96	43.5	2471.2
WOB(ton)	4866	6.637	0.01	20.72
RPM(rpm)	4866	85.11	20.0	135.0
TORQUE(lb*ft)	4866	4911.23	664.0	10077.0
SPP(psi)	4866	1413.143	217.0	2144.0
FLWpumps(l/mn)	4866	2608.83	1225.0	4039.0
MW(gm/cc)	4866	1.112	1.04	1.21
Pump(spm)	4866	70.52	14.0	98.0
ECD(gm/cc)	4866	1.142	1.01	1.29
BIT SIZE(inch)	4866	12.534	8.5	17.5
ROP(m/hr)	4866	15	1.31	81.55

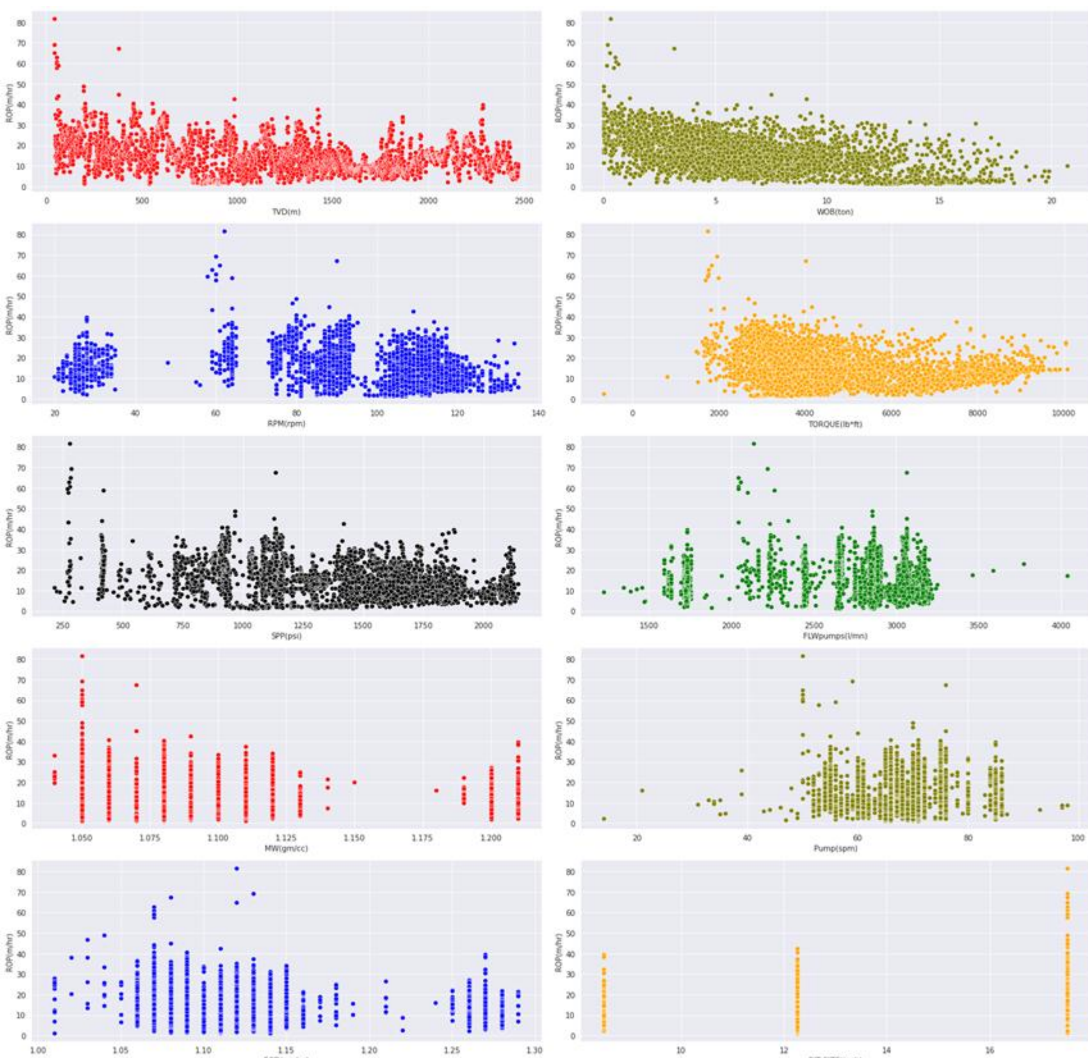

Figure 4: Scatterplots ROP Vs input parameters.

2.1. Model development

Gradient Boosting is a popular machine learning algorithm that can be used for regression tasks. In Python, we can import the Gradient Boosting Regressor library from the Scikit-Learn package to build and train our model. The Gradient Boosting Regressor is an ensemble model that combines the predictions of multiple decision trees in order to improve accuracy.

Gradient Boosting Regressor (GBR) is a powerful machine learning algorithm used for regression problems. It has several hyperparameters that can be tuned to optimize the model's performance. The number of estimators refers to the number of trees in the forest, and increasing this hyperparameter can improve the model's accuracy but can also lead to overfitting. The learning rate determines the size of the step taken during each sequential iteration, affecting the convergence rate and the model's ability to generalize. Choosing the appropriate loss function, such as mean squared error or mean absolute error, can also enhance the model's performance by minimizing the errors between predicted and actual values.

In addition to these hyperparameters, there is also the criterion hyperparameter which is used to measure the quality of a split. The default criterion in the Scikit-Learn Gradient Boosting Regressor is the mean squared error, but other criteria such as the mean absolute error or the Huber loss can also be used.

To obtain the best performance from the Gradient Boosting Regressor, we can use techniques such as grid search or trial and error to find the optimal hyperparameters. For example, we can use grid search to test different combinations of hyperparameters and evaluate the performance of each combination using cross-validation. Table 2 shows an example of the best hyperparameters that can be used to build a Gradient Boosting Regressor model:

Table 2: The best hyperparameters that used in study.

Hyperparameters	**The best choice**	**Description**
Loss	Squared error	Optimization function
Learning rate	0.05	learning rate ranges from 0 to 1 and represents the rate of adjustment of weights and the speed of learning.
n estimators	100	The number of trees
criterion	MSE	The function for determining a split's quality.
Min. samples split	4	To split an internal node, the minimal number of samples is necessary.
Min. samples leaf	4	The bare minimum of samples that must be present at a leaf node.
Max. depth	None	The tree's greatest depth.

2. Results and Discussion

After completing data preparation and feature engineering, our model was trained using 10 scaled predictors including TVD(m), WOB(ton), RPM(rpm), TORQUE(lb*ft), SPP(psi), FLWpumps(l/mn), MW(gm/cc), Pump(spm), ECD(gm/cc), BIT SIZE(inch), and ROP(m/hr). Our dataset consisted of 4866 data samples, as shown in Table 1.

To ensure accurate model performance, we split our data into a 70% training set and a 30% test set. Next, we optimized the model's hyperparameters, including the gradient boosting method's hyperparameters, using GridSearchCV in Python. This approach allowed us to select the optimal values for the hyperparameters and improve the model's predictive power.

Once the hyperparameters of a model are optimized, the next step is to train and test it using a certain percentage of the available data. In this case, 70% of the data was used for training, and the remaining 30% for testing. The obtained R2 values of 0.9947 and 0.8611 for training and testing, respectively, indicate that the model has acceptable performance. Interestingly, the model is more accurate in predicting ROP in the lower range, indicating better performance in deeper formations where ROP is typically higher. These results demonstrate the effectiveness of the model in predicting ROP.

Figure (7) illustrates the successful implementation of a trained model to estimate the rate of penetration for a new well. This marks a significant milestone in the field of well drilling optimization.

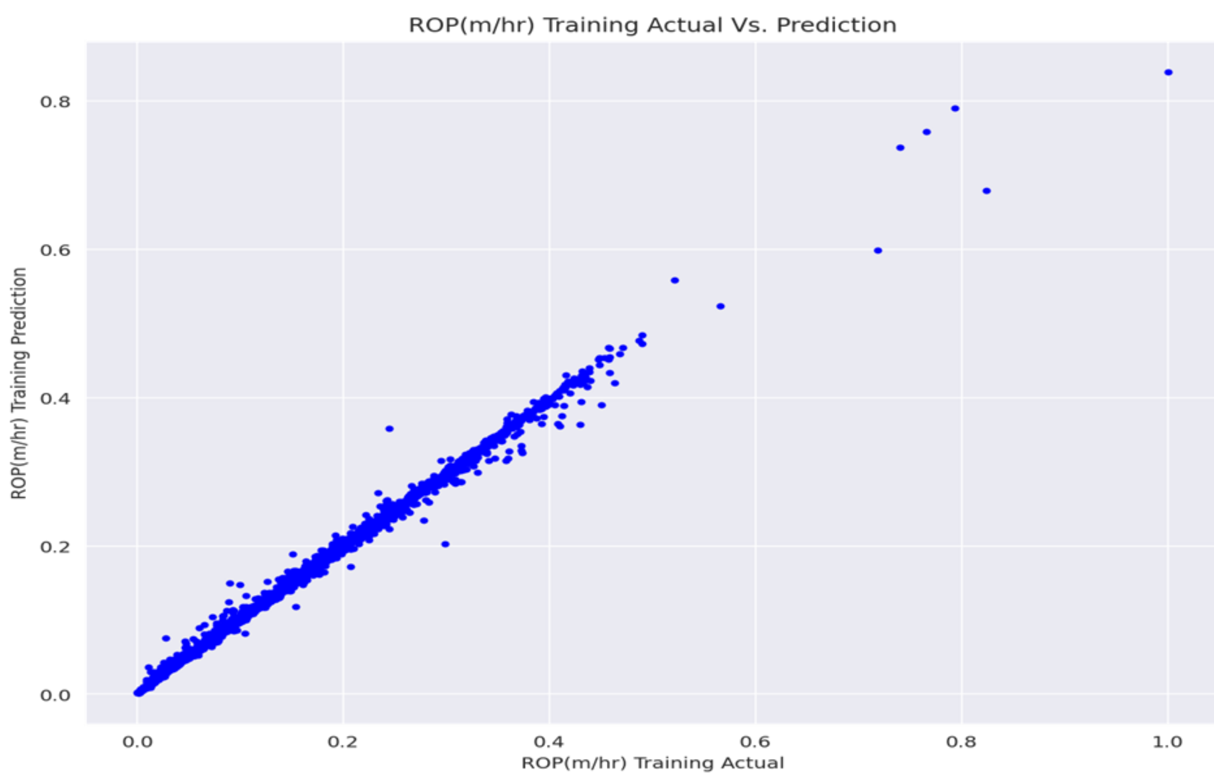

Figure 5: Actual and predicted ROP for training data.

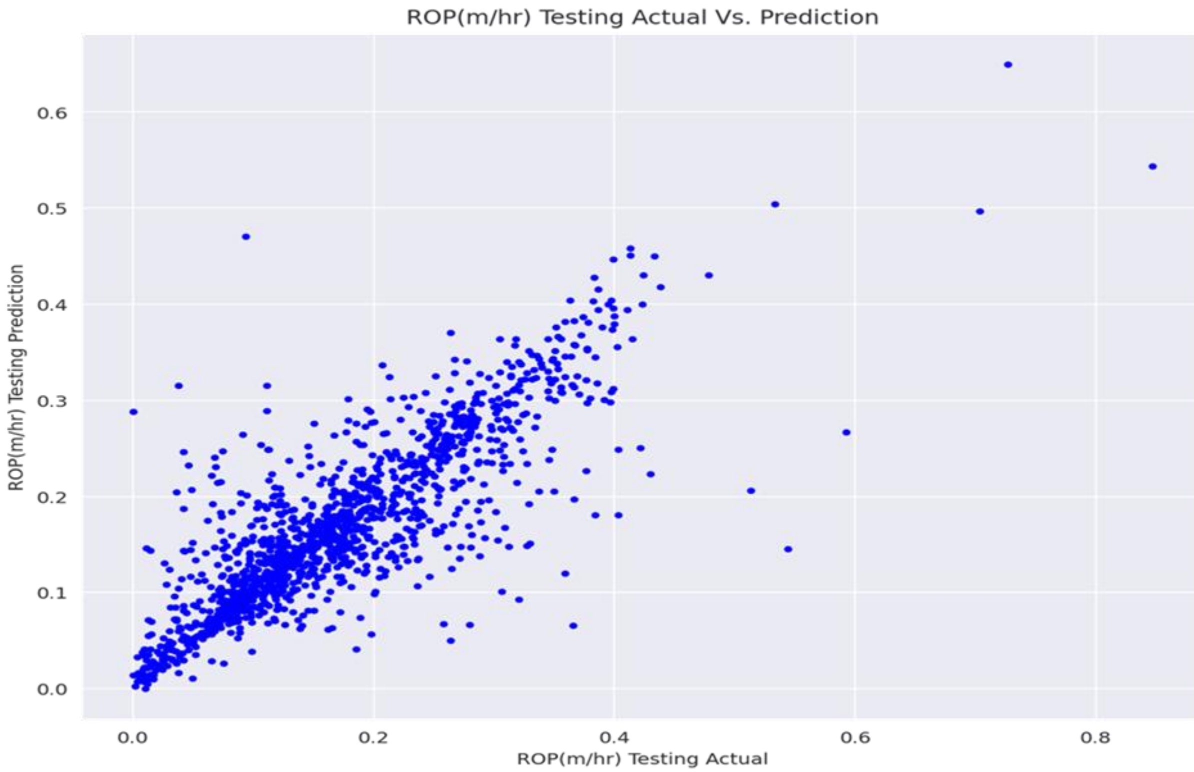

Figure 6: Actual and predicted ROP for testing data.

Figure 7: Actual and predicted ROP for the new well.

3. Conclusion

The following conclusions may be made based on the findings from the previous section:

The Gradient Boosting (GB) model was developed to forecast the Rate of Penetration (ROP) in Rumaila oilfield based on drilling operation parameters and drilling mud properties. The model was trained and tested on two wells, with one well used for implementation. The GB model achieved high accuracy in predicting ROP, especially in deeper formations, with an R2 value of 0.8611. Overall, the model demonstrated significant potential for optimizing drilling operations in the oil and gas industry.

References

[1] Riazi, Mohsen, Hossein Mehrjoo, Reza Nakhaei, Hossein Jalalifar, Mohammadhadi Shateri, Masoud Riazi, Mehdi Ostadhassan, and Abdolhossein Hemmati-Sarapardeh. "Modelling rate of penetration in drilling operations using RBF, MLP, LSSVM, and DT models." Scientific Reports 12, no. 1 (2022): 11650.doi: https://doi.org/10.1038/s41598-022-14710-z.

[2] Bahari, Mohamad Hasan, Abouzar Bahari, and Hamidreza Moradi. "Intelligent drilling rate predictor." International Journal of Innovative Computing, Information and Control 7, no. 2 (2011): 1511-1520.

[3] Hadi, Hassan Abdul. "Correlation of penetration rate with drilling parameters for an Iraqi field using mud logging data." Iraqi Journal of Chemical and Petroleum Engineering 16, no. 3 (2015): 35-44.

[4] Kaiser, Mark J. "A survey of drilling cost and complexity estimation models." International Journal of Petroleum Science and Technology 1, no. 1 (2007): 1-22.

[5] Metropolis, Nicholas, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. "Equation of state calculations by fast computing machines." The journal of chemical physics 21, no. 6 (1953): 1087-1092.

[6] Eskandarian, Sajad, Peyman Bahrami, and Pezhman Kazemi. "A comprehensive data mining approach to estimate the rate of penetration: Application of neural network, rule based models and feature ranking." Journal of Petroleum Science and Engineering 156 (2017): 605-615.doi: https://doi.org/10.1016/j.petrol.2017.06.039.

[7] Ahmed, Omogbolahan S., Ahmed A. Adeniran, and Ariffin Samsuri. "Computational intelligence-based prediction of drilling rate of penetration: A comparative study." Journal of Petroleum Science and Engineering 172 (2019): 1-12.doi: https://doi.org/10.1016/j.petrol.2018.09.027.

[8] Soares, Cesar, and Kenneth Gray. "Real-time predictive capabilities of analytical and machine learning rate of penetration (ROP) models." Journal of Petroleum Science and Engineering 172 (2019): 934-959.doi: https://doi.org/10.1016/j.petrol.2018.08.083.

[9] Ataei, Mohammad, Reza KaKaie, Mehdi Ghavidel, and Omid Saeidi. "Drilling rate prediction of an open pit mine using the rock mass drillability index." International Journal of Rock Mechanics and Mining Sciences 73 (2015): 130-138.doi: https://doi.org/10.1016/j.ijrmms.2014.08.006.

[10] Akgun, Ferda. "Drilling rate at the technical limit." International Journal of Petroleum Science and Technology 1, no. 1 (2007): 99-119.

[11] Mendes, José Ricardo P., Tiago C. Fonseca, and Adriane Serapião. "Applying a genetic neuro-model reference adaptive controller in drilling optimization." World oil (2007): 29-36.doi: https://orcid.org/0000-0001-9728-7092.

[12] Arabjamaloei, R., and S. Shadizadeh. "Modeling and optimizing rate of penetration using intelligent systems in an Iranian southern oil field (ahwaz oil field)." Petroleum Science and Technology 29, no. 16 (2011): 1637-1648.doi: https://doi.org/10.1080/10916460902882818.

[13] Al-Jaberi, Mohanad H., and Hayder K. Al-Mayyahi. "Wireline logging response and true core analysis of the upper shale member of Zubair Formation, Rumaila oilfield, Southern Iraq." The Iraqi Geological Journal (2018): 28-40.

[14] Al-Ansari, R. 1993. The Petroleum Geology of the Upper Sandstone Member of the Zubair Formation in the Rumaila South. Geological Study, Ministry of Oil, Baghdad, Iraq

[15] Al-Rawi, Munim. "Rumaila-West Qurna: A Unique Supergiant Field." (2018).

[16] Handhal, Amna M., Alaa M. Al-Abadi, Hussein E. Chafeet, and Maher J. Ismail. "Prediction of total organic carbon at Rumaila oil field, Southern Iraq using conventional well logs and machine learning algorithms." Marine and Petroleum Geology 116 (2020): 104347.doi: https://doi.org/10.1016/j.marpetgeo.2020.104347.

[17] Ahmed, Omogbolahan S., Ahmed A. Adeniran, and Ariffin Samsuri. "Computational intelligence based prediction of drilling rate of penetration: A comparative study." Journal of Petroleum Science and Engineering 172 (2019): 1-12. https://doi.org/10.1016/j.petrol.2018.09.027

[18] Shi, Xian, Gang Liu, Xiaoling Gong, Jialin Zhang, Jian Wang, and Hongning Zhang. "An efficient approach for real-time prediction of rate of penetration in offshore drilling." Mathematical Problems in Engineering 2016 (2016).doi https://doi.org/10.1155/2016/3575380.

[19] Azar, Hadi Fathipour, Timo Saksala, and Seyed-Mohammad Esmaiel Jalali. "Artificial neural networks models for rate of penetration prediction in rock drilling." Rakenteiden Mekaniikka 50, no. 3 (2017): 252-255.doi: https://doi.org/10.23998/rm.64969.

[20] Hegde, Chiranth, and Ken Gray. "Evaluation of coupled machine learning models for drilling optimization." Journal of Natural Gas Science and Engineering 56 (2018): 397-407.doi: https://doi.org/10.1016/j.jngse.2018.06.006.

[21] Bodaghi, Asadollah, Hamid Reza Ansari, and Mahsa Gholami. "Optimized support vector regression for drillingrate of penetration estimation." Open Geosciences 7, no. 1 (2015).doi: https://doi.org/10.1515/geo-2015-0054.

[22] Hegde, Chiranth, Hugh Daigle, Harry Millwater, and Ken Gray. "Analysis of rate of penetration (ROP) prediction in drilling using physics-based and data-driven models." Journal of petroleum science and Engineering 159 (2017): 295-306.doi: https://doi.org/10.1016/j.petrol.2017.09.020.

[23] Hegde, Chiranth, Scott Wallace, and Ken Gray. "Using trees, bagging, and random forests to predict rate of penetration during drilling." In SPE Middle East Intelligent Oil and Gas Conference and Exhibition. OnePetro, 2015.doi: https://doi.org/10.2118/176792-MS.

[24] Ashrafi, Seyed Babak, Mohammad Anemangely, Mohammad Sabah, and Mohammad Javad Ameri. "Application of hybrid artificial neural networks for predicting rate of penetration (ROP): A case study from Marun oil field." Journal of petroleum science and engineering 175 (2019): 604-623.doi: https://doi.org/10.1016/j.petrol.2018.12.01.

[25] Jia, Weikuan, Dean Zhao, and Ling Ding. "An optimized RBF neural network algorithm based on partial least squares and genetic algorithm for classification of small sample." Applied Soft Computing 48 (2016): 373-384.doi: https://doi.org/10.1016/j.asoc.2016.07.037.

[26] Mahdi, Doaa Saleh. "Predicting Drilling Rate of Penetration Using Artificial Neural Networks." In IOP Conference Series: Materials Science and Engineering, vol. 1067, no. 1, p. 012150. IOP Publishing, 2021.doi: https://doi.org/10.1088/1757-899X/1067/1/012150.

[27] Zhang, Haowei, Junwei Xie, Jiaang Ge, Wenlong Lu, and Binfeng Zong. "An entropy-based PSO for DAR task scheduling problem." Applied soft computing 73 (2018): 862-873.

[28] Kennedy, James, and Russell Eberhart. "Particle swarm optimization." In Proceedings of ICNN'95-international conference on neural networks, vol. 4, pp. 1942-1948. IEEE, 1995.

[29] Breiman, Leo. "Bagging predictors." Machine learning 24 (1996): 123-140.

[30] Zhou, Ligang, Hamido Fujita, Hao Ding, and Rui Ma. "Credit risk modeling on data with two timestamps in peer-to-peer lending by gradient boosting." Applied Soft Computing 110 (2021): 107672.doi: https://doi.org/10.1016/j.asoc.2021.107672.

[31] Wu, Changxing, Chaowen Hu, Ruochen Li, Hongyu Lin, and Jinsong Su. "Hierarchical multi-task learning with CRF for implicit discourse relation recognition." Knowledge-Based Systems 195 (2020): 105637.doi: https://doi.org/10.1016/j.knosys.2020.105637.

[32] Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785-794. 2016.doi: https://doi.org/10.1145/2939672.2939785.

[33] Burlina, Philippe M., Neil Joshi, Michael Pekala, Katia D. Pacheco, David E. Freund, and Neil M. Bressler. "Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks." JAMA ophthalmology 135, no. 11 (2017): 1170-1176. doi: https://doi.org/10.1001/jamaophthalmol.2017.3782.

[34] A. K. Salih and H. A. Abdul Hussein, "Lost circulation prediction using decision tree, random forest, and extra trees algorithms for an Iraqi oil field," Iraqi Geological Journal, vol. 55, no. 2E, pp. 111–127, 2022.

[35] Drucker, Harris, and Corinna Cortes. "Boosting decision trees." Advances in neural information processing systems 8 (1995).

[36] Friedman, Jerome H. "Greedy function approximation: a gradient boosting machine." Annals of statistics (2001): 1189-1232.

[37] Ji, Chuanyi, and Sheng Ma. "Combinations of weak classifiers." Advances in Neural Information Processing Systems 9 (1996).

[38] Twenge, Jean M., Thomas E. Joiner, Megan L. Rogers, and Gabrielle N. Martin. "Increases in depressive symptoms, suicide-related outcomes, and suicide rates among US adolescents after 2010 and links to increased new media screen time." Clinical Psychological Science 6, no. 1 (2018): 3-17.doi: https://doi.org/10.1177/2167702617723376.

[39] A. K. Salih, H. A. Abdul Hussein, and S. H. Hamza, "Artificial neural networks to predict lost circulation zones at southern Iraq oilfield," Journal of Petroleum Research and Studies, vol. 13, no. 1, pp. 16–31, 2023.

[40] Baturynska, Ivanna, and Kristian Martinsen. "Prediction of geometry deviations in additive manufactured parts: comparison of linear regression with machine learning algorithms." Journal of Intelligent Manufacturing 32 (2021): 179-200.