Journal of Mining and Environment (JME), Vol. 14, No. 3, 2023, 755-770

Shahrood University of Technology

Journal of Mining and Environment (JME)

Journal homepage: www.jme.shahroodut.ac.ir

Iranian Society of Mining Engineering (IRSME)

Rate of Penetration Prediction in Drilling Operation in Oil and Gas Wells by K-nearest Neighbors and Multi-layer Perceptron Algorithms

Yahia El-Sayed *, Adel M.Salem, and Shady El-Rammah

Department of Petroleum Engineering, Faculty of Petroleum and Mining Engineering, Suez University, Suez, Egypt

Article Info

Received 6 February 2023

Received in Revised form 7 May 2023

Accepted 1 June 2023

Published online 1 June 2023

DOI:10.22044/jme.2023.12694.2306

Keywords

Rate of penetration

Machine learning

Drilling, Multi-layer perceptron

K-nearest neighbors

Abstract

The rate of penetration plays a key role in maximizing drilling efficiency, so it is essential for the drilling process optimization and management. Traditional mathematical models have been used with some success to predict the rate of penetration in drilling. Due to the high complexity and non-linear behavior of drilling parameters with the rate of penetration, these mathematical models cannot accurately and comprehensively predict the rate of penetration. Machine learning (ML) seems to be an attractive alternative to model this complicated physical process. This research paper introduces new data-driven models used to predict ROP using different parameters such as (depth, weight on bit (WOB), revolution per minute (RPM), Torque (T), standpipe pressure (SPP), flow in pump (pumping flow rate(Q), mud weight, hours on bit (HOB), revolutions on bit, bit diameter, total flow area (TFA), pore pressure, overburden pressure, and pit volume). Data-driven models are built using two different machine learning techniques, using 1771 raw real field data. The coding is built using the python programming language. The k-nearest neighbors (KNN) model predicting ROP for the training dataset show a correlation coefficient (R2) of 0.94. The multi-layer perceptron (MLP) model predicting ROP for the training dataset show a correlation coefficient (R2) of 0.98. We can conclude that MLP has a better accuracy, and removing outliers enhances model performance.

1. Introduction

The drilling process is the foundation of the oil and gas business. Drilling performance needs to be regularly tracked to save drilling costs and enhance oilfield operations. There has been a lot of work done to optimize the drilling process and prevent drilling problems. Drilling speed and expense are typically connected. To successfully conduct a drilling operation, and as a result, reduce drilling expenses; a suitable rate of penetration (ROP) must be established [1]. The minimum cost per foot of drilled well does not always follow from the highest on-bottom drilling rate. Costs could be increased by additional factors such as rapid bit wear and equipment breakdown [2].

1.1. Problem statement

In the past, various drilling models were presented to describe the impact of drilling variables, environment, and geology on the ROP. **Figure 1** shows some of the traditional mathematical models used to predict the rate of penetration in drilling and the factors included in each model. These models predicted ROP with some effectiveness such as Speer's model [3], cunningham's model [4], Bingham's model [5], Bourgoyne and Young's model (BYM) [6], and Harreland and Rampersad's model [7].

These mathematical methods above cannot completely and precisely anticipate the rate of penetration due to the high complexity and non-linear behavior of drilling variables regarding the rate of penetration. Given the development of both software and hardware tools along with the appearance of the issues and limitations of earlier mathematical models, machine learning seems to be a compelling option to represent this complex physical process.

Corresponding author: Yahia.Khamis@pme.suezuni.edu.eg (Y. El-Sayed)

This study aims to use machine learning different techniques to develop a model that predicts the rate of penetration using field data from mud logs and petro-physical data. In addition, comparing the output data of these models with the actual data to show the power of computational programs over traditional methods in solving problems and predicting variables that are hard to be modeled mathematically.

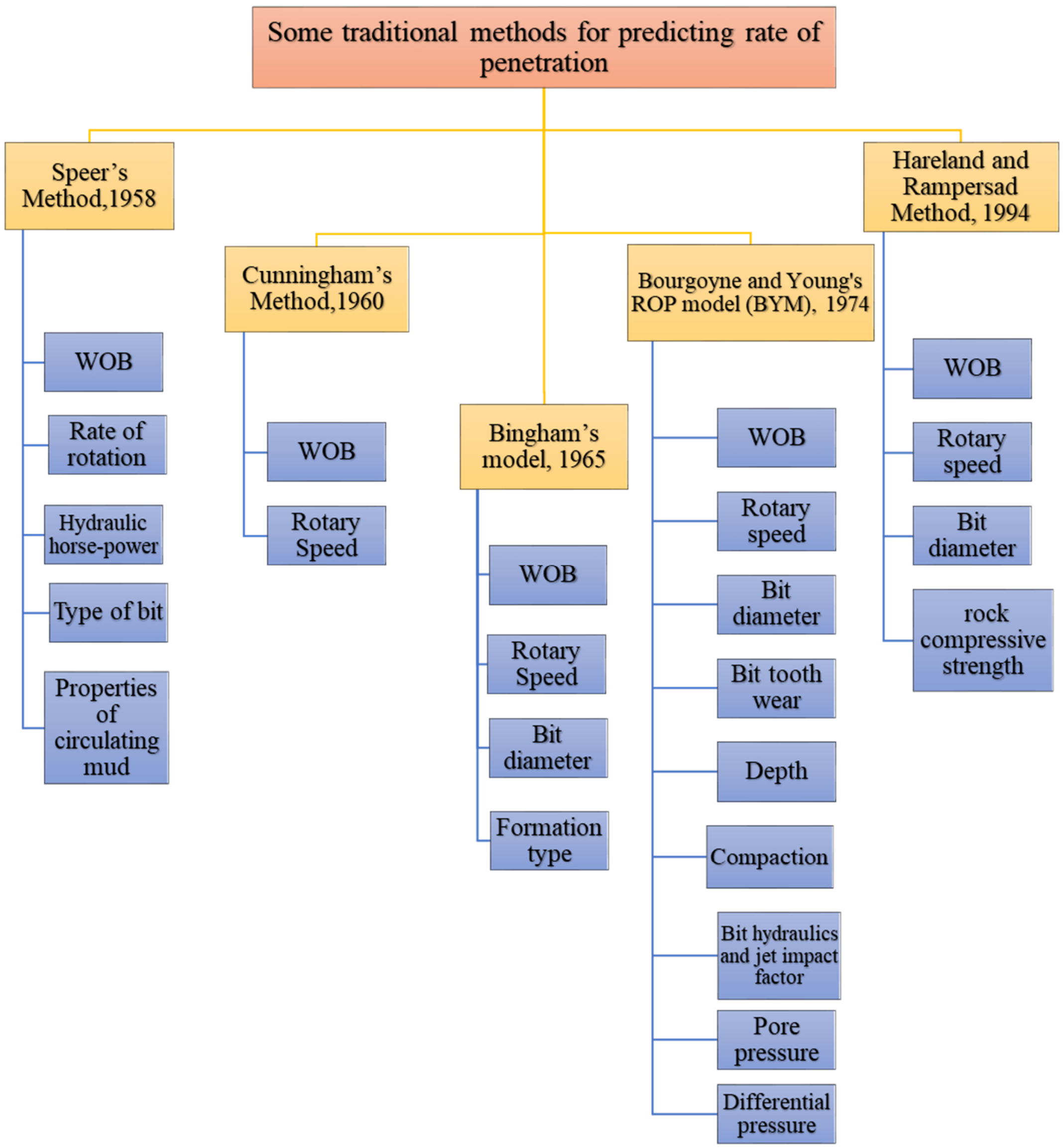

Figure 1. Some of the traditional mathematical models used to predict ROP.

2. Literature Review

2.1. Rate of penetration

The depth of penetration obtained per unit of time is known as the penetration rate or drilling speed [8]. ROP is usually reported in ft/h (field units) or m/h (SI units). In any engineering study of rotary drilling, it is convenient to divide the factors that affect the rate of penetration into the following list (rig and personal efficiency, mechanical factors, hydraulic factors, formation characteristics, mud properties, and bit type), as shown in **Figure 2**. Because of the complex and non-linear relationships between each of these characteristics and the ROP as well as the innate inaccuracies in data interpretation, it has been difficult to demonstrate a comprehensive relation [9].

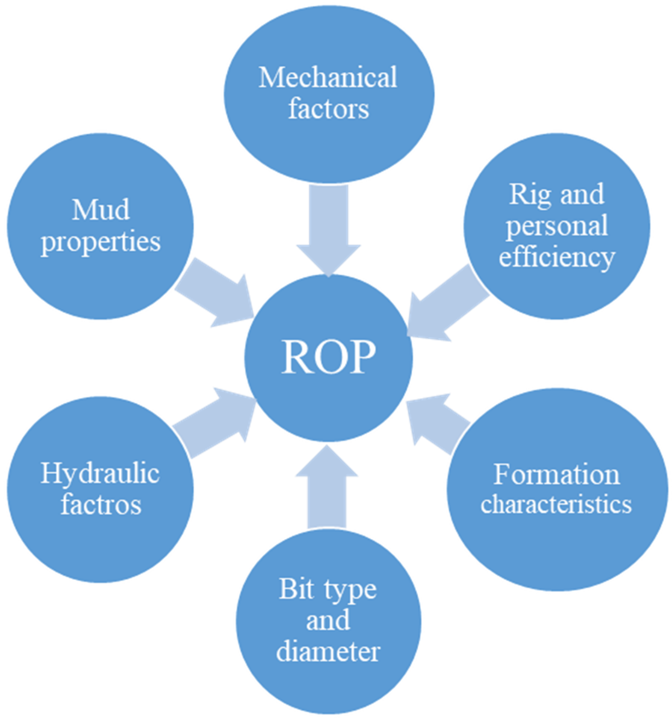

Figure 2. Factors affecting rate of penetration.

2.1.1. Mechanical factors

2.1.1.1. Weight on bit

The desired weight on bit is achieved by using a number of drill collars. Due to the frictional drag of the drill string on the borehole wall, actual downhole weight on bit values are usually lower than surface WOB values, according to field evidence. This is particularly true in the case of directional wells [10].

The effect of weight on bit on penetration rate can be graphically represented by a three-part curve (**Figure 3**). The increase in WOB from 0 to w1 corresponds to section ab of the curve. The contact pressure between the bit operating parts and the rock in this region is lower than the rock's hardness. As a result, bit operating parts (such as the teeth of a roller cone bit) cannot penetrate the rock and smash it. Only friction between the bit and the bottom causes rock disintegration. Because the normal pressure (or load) is exactly proportional to the friction force, increasing the weight on the bit increases the penetration rate.

The contact pressure is still lower than rock hardness in the region bc of the curve but the rate of penetration develops quicker than the weight on bit. The effect of fatigue phenomenon is frequently used to describe such a relation between two elements. When drilling at a weight on bit WOB < W2, the impact of a bit tooth is not enough to smash the rock but it may be enough to fracture it. When a bit tooth hits the same area on the bottom after a few revolutions, new fractures may form, and old fractures could become deeper. The rock grows weaker as a result of many hits, and a future collision may produce rock crushing. When the weight on bit increases from w1 to w2, fewer and fewer impacts are required to cause rock fatigue failure. For this reason, the penetration rate grows faster than the weight on bit. Rock mass drillability depends on a number of parameters including intact rock/rock mass properties, machine specifications, and operational parameters [11].

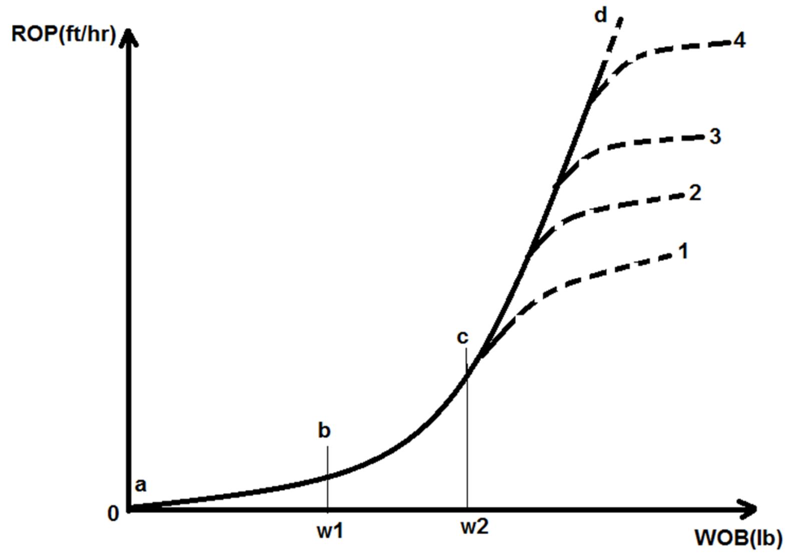

Figure 3. Effect of weight on bit on penetration rate, *modified after* [12].

The contact pressure is larger than the rock hardness in the region cd of the curve, and the rate of penetration returns to being a linear relationship of the weight on bit. Only if proper bottom cleaning is performed, can such a linear relationship exist. If bottom cleaning isn't done properly, the relationship may depart from the straight line cd and go in one of two directions (curves 1, 2, 3 or 4). The linear relationship will be extended towards higher bit loads when the circulation rate is increased.

It is clear from this discussion that the meterage per bit will likely reach a peak at a particular weight on bit. Field practice has proven the existence of such a limit in some cases, and the prospect of lowering the meterage per bit at large bit weights should not be underestimated [13].

2.1.1.2. Rotary speed

An increase in rotational speed has some impacts on bit performance. When the rotational speed is raised, it causes an increase in the number of tooth impacts, an increase in the speed of impacts, and a decrease in the time the teeth are in contact with the drilled rock.

Every tooth hit causes a small amount of rock to disintegrate. The amount of rock disintegrated increases as the frequency of strikes increases, resulting in a commensurate increase in the penetration rate. Increased impact speed adds a bigger volume of rock with each hit, resulting in a higher penetration rate.

Numerous parameters such as the origin of rocks formation, Mohs hardness, texture of rock (shape and size of rock grains), porosity, density, abrasiveness, rigidity, P-wave velocity, elasticity and plasticity, UCS, tensile strength, affect the drilling rate and drillability of rocks [14].

The energy of the hit is used for rock breakup and shattering while a bit tooth is in contact with the rock. Rock deformation is a gradual process that takes time to complete. The duration of contact between a bit teeth and the rock should be long enough for the deformation to develop to its fullest extent for optimal impact efficiency. When the rotational speed increases, the time of contact reduces, and the effectiveness of each hit diminishes. With increasing rotary speed, the overall impact of the components involved increases the penetration rate. **Figure 4** shows the response of rate of penetration with changes in rotation speed.

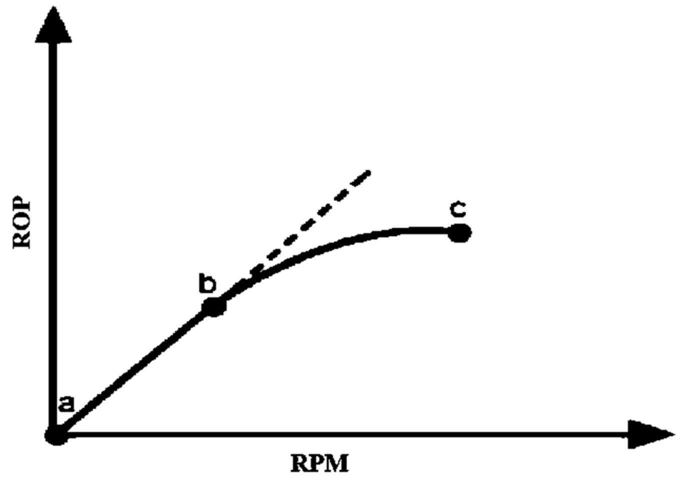

Figure 4. Typical responses of ROP for changing rotation speeds, *modified after* [15].

2.2. Machine learning

Machine learning is the study of computer algorithms that provide systems the ability to automatically learn from experience [16]. It's generally regarded as an artificial intelligence subfield. Computers are now capable of autonomous decision-making thanks to machine learning algorithms. Such decisions are achieved by locating significant underlying patterns in vast, intricate data sets [17].

Machine learning systems can be categorized into:-

- Supervised learning
- Unsupervised learning
- Semi-supervised learning
- Reinforcement learning

2.2.1. K-nearest neighbors

Machine learning models can be divided into parametric and nonparametric models. Parametric models, in general, have some parameters that are directly learned from data. Non-parametric models, in contrast, do not contain any parameters that may be learned from data [18].

K-nearest neighbors (KNN) is a non-parametric algorithm based on the similarity measure (e.g. distance functions such as Euclidian distance between two samples or hamming distance between the binary vectors). This non-parametric pattern categorization method was first introduced by Fix and Hodges in 1951, and is currently known as the k-nearest neighbors method [19].

KNN retains all training samples in memory, in contrast to other learning methods that permit dumping the training data once the model is constructed [20]. The algorithm finds the k nearest neighboring points in the dataset using the similarity measure and then forecasts the output category based on the most frequent class within these k neighbors or the average label in case of regression [21]. **Figure 5** shows an example of a regression problem using KNN.

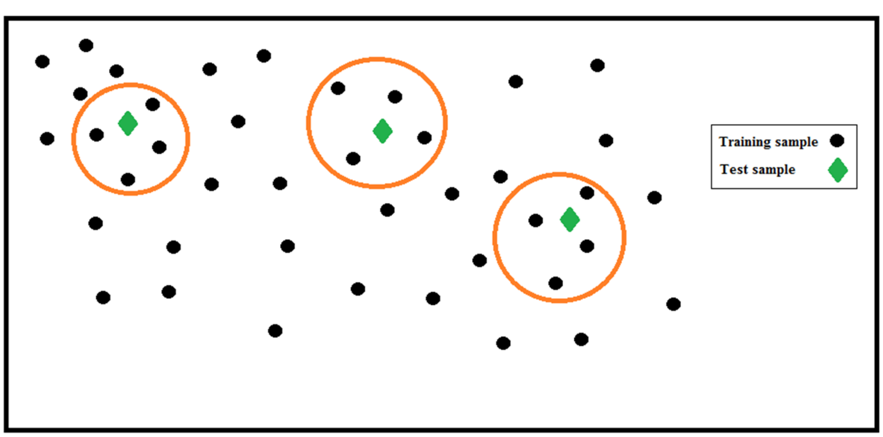

Figure 5. Regression and missing value imputation using the KNN method with a fixed k value, that is k = 4.

2.2.2. Multi-layer perceptron (MLP)

The operation of biological neurons inside the human brain serves as inspiration for artificial neural networks (ANN). An ANN's key benefits include its capacity to investigate exceedingly intricate nonlinear correlations between the variables [22]. The most fundamental type of deep learning algorithm is a multi-layer perceptron (MLP). Deep neural networks' fundamental building components are neurons. To carry out non-linear transformations, the neurons may employ a variety of activation functions. Among the commonly utilized activation functions are the sigmoid function and the hyperbolic tangent function. The rectified linear unit is one of the most advanced activation functions (ReLU) used [21].

A multi-layer perceptron (MLP) has an input layer, one to several hidden layers and an output layer, as shown in **Figure 6**. MLP is a common feedforward ANN architecture. The quantity of input and output variables, respectively, determines the quantity of input and output neurons [23]. The number of hidden layers and also the number of neurons in each hidden layer can be chosen optimally. Each neuron multiplies inputs by a weight and adds them with a value termed bias. Then, a function known as the transfer function is used to transform their summation (activation function) [24].

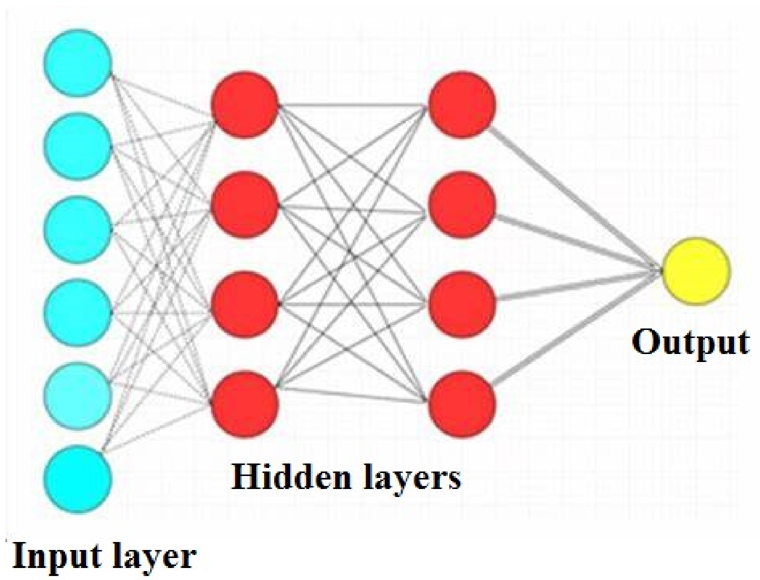

Figure 6. Schematic view of the MLP model.

Engineering in general, and petroleum engineering in particular, can use machine learning in a variety of ways. In order to create models for use in reservoir, production, and drilling engineering, machine learning techniques have been applied. The following table (**Table 1**) lists some of the models created to forecast rate of penetration, summarising the methods employed, the features that each model covered, and its level of accuracy.

3. Methods and Materials

Figure 7 shows the flow chart of steps that would be followed until the models are built.

3.1. Collecting data

In this study, real field data from an offshore vertical gas well (Well-X) in the Middle East's Offshore Nile Delta were used to improve machine learning models for the ROP using inputs of depth, WOB, RPM, T, SPP, Q, mud weight, HOB, revolutions on bit, bit diameter (D_{bit}), TFA, pore pressure (PP), overburden pressure (OVB), and pit volume.

The pressure data were used from offset wells having the same stratigraphic column. The data was recorded by the surface real-time data transmitter sensor. Through 17½ "Hole and 12¼" Hole sections, 1771 data points between depths of 1207 m and 2092 m were gathered. Identical two bits having the same IADC code were used to drill both parts but the two bits have different diameters and nozzle sizes. **Table 2** shows the summary of the statistical parameters of the data.

Table 1. Summary of some of the data-driven models used to predict ROP.

Model	ML algorithms	Parameters	Number of data points used to build the model	Accuracy MAPE	Accuracy RMSE	Accuracy R^2
(Hazbeh *et al*., 2021) [25]	MLP	WOB, RPM, SPP, T, Q, ECD, BPL	1878	--	--	0.99
(Anemangely *et al*., 2020) [26]	MLP	D, WOB, RPM, mud weight, Q, SPP, porosity log, density log, gamma log, compressional and shear wave slowness	--	--	0.064	0.921
(Sabah *et al*., 2019) [27]	MLP	Neutron porosity density, shear wave velocity, compressional wave velocity, gamma ray, WOB, bit rotational speed (BRS), pump pressure, bit flow rate (BFR), mud weight, pore pressure gradient	1000	--	1.15	0.928
(Moran *et al*., 2010) [28]	ANN	Rock strength, rock type, abrasion, WOB, RPM, mud weight	--	--	--	0.8
(Elkatatny, 2018) [29]	ANN	WOB, SPP, RPM, Q, T, M.wt, FV, PV	3333	--	--	0.99

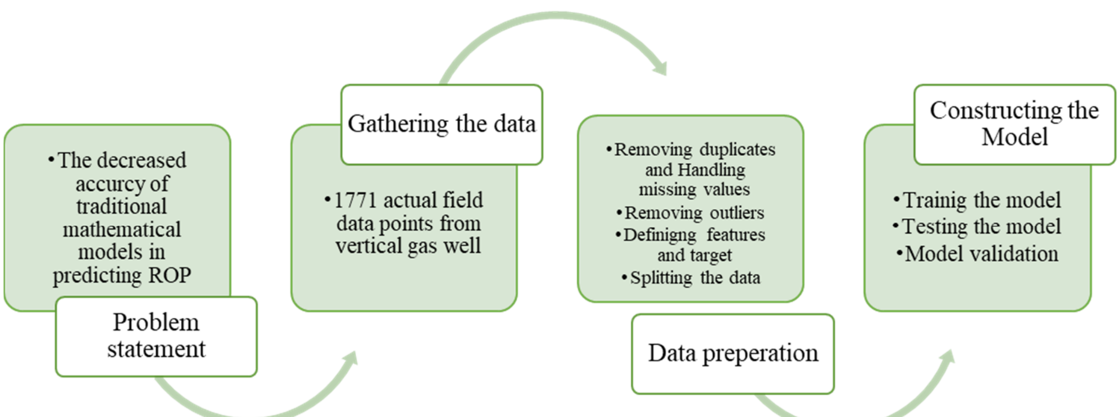

Figure 7. Steps until developing the model.

Table 2. Statistical summary of the data.

Statistical parameter	ROP (m/hr)	WOB (klb)	RPM (rpm)	T (ft-lb)	SPP psig	Q (gpm)	Mwt (ppg)	HOB (hr)	Rev On Bit (krev)	D_{BIT} (in)	TFA in^2	pit volume (bbl)	PP (ppg)	OVB (ppg)
Maximum	25.3	32.8	121	4948.7	3429	1079	10.72	43.85	304	17.50	1.17	785.26	10.32	16.04
Minimum	3.7	0.6	49	1749.7	1645	508	10.20	0.05	0.1	12.25	0.99	396.82	8.85	12.51
Range	21.6	32.2	72	3199	1784	571	0.52	43.80	303.9	5.25	0.17	388.44	1.47	3.53
Mean	12.35	8.45	115.28	2666.47	2908.09	960.78	10.50	19.47	128.27	14.91	1.08	698.69	9.73	14.5
Median	11.60	7.6	119	2623	2952	940	10.50	18.65	122.1	17.50	1.17	695.85	9.81	14.59
Standard Deviation	3.61	4.88	14.51	484.33	291.67	60.73	0.13	11.61	82.94	2.63	0.09	38.50	0.36	1.04

3.2. Data preparation

There are no duplicates in the data, and no values were missing. An observation that differs so significantly from previous observations that it raises the possibility that it was produced by a distinct mechanism is referred to as an outlier [30].

There are several methods for finding and removing outliers but isolation forest was the one employed in this study. Outliers in 14.7% of the data were found and eliminated. **Figure 8** and **Figure 9** show SPP before and after outliers were removed, using a box-whisker plot to show outliers.

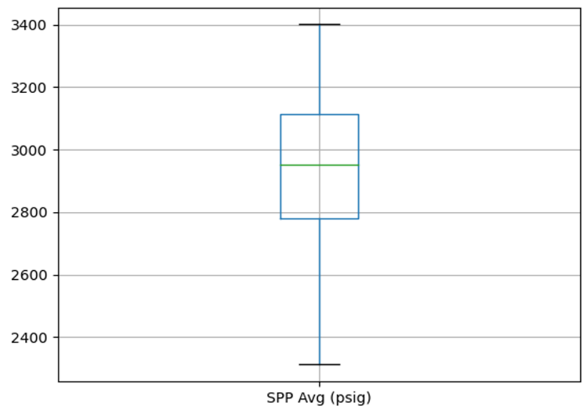

Figure 8. SPP after removing outliers.

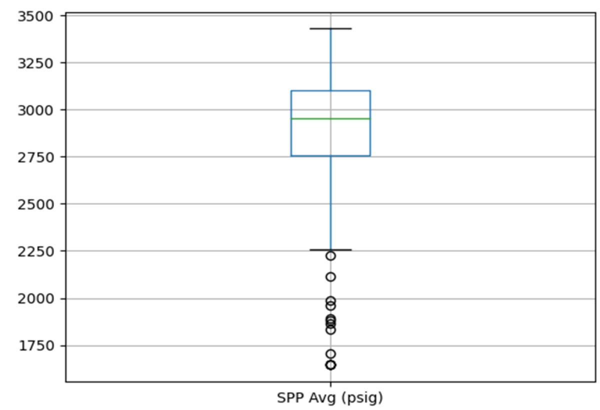

Figure 9. SPP before removing outliers.

The data quality is improved by eliminating outliers. It improved the coefficients of correlation between the features and the target. **Figure 10** displays the correlation coefficients that were calculated from the correlation matrix before the outliers were removed. **Figure 11** shows a comparison between correlation coefficients before and after removing outliers. **Figure 12** shows the data points of pumping rate before removing outliers and after removing outliers and points that have been removed.

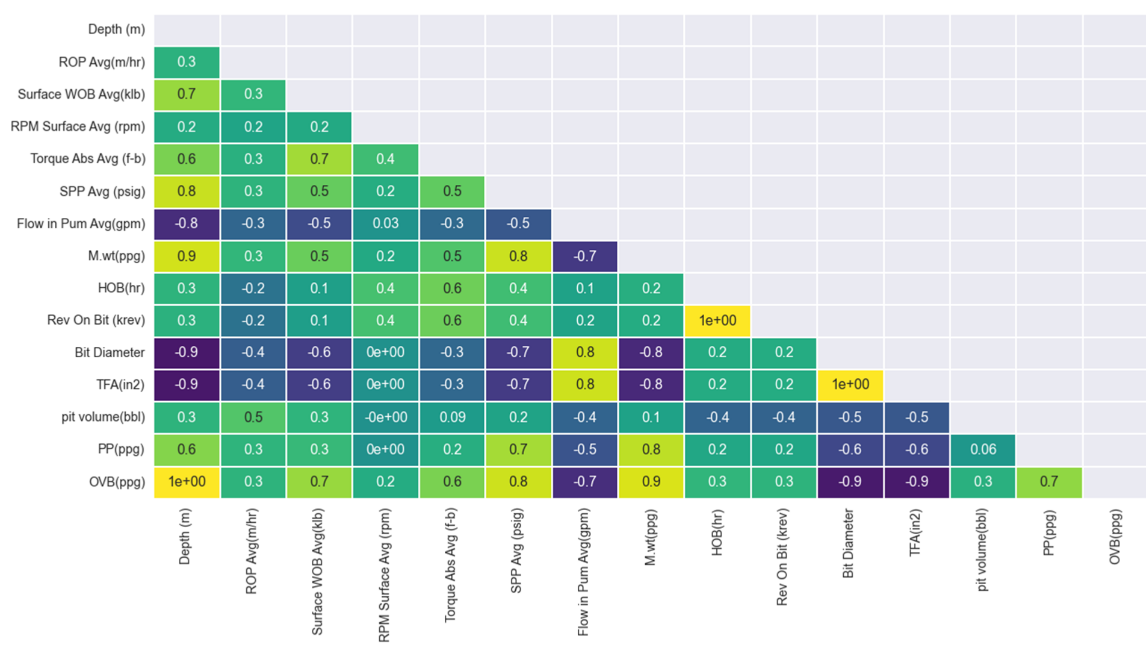

Figure 8. Correlation matrix before removing outliers.

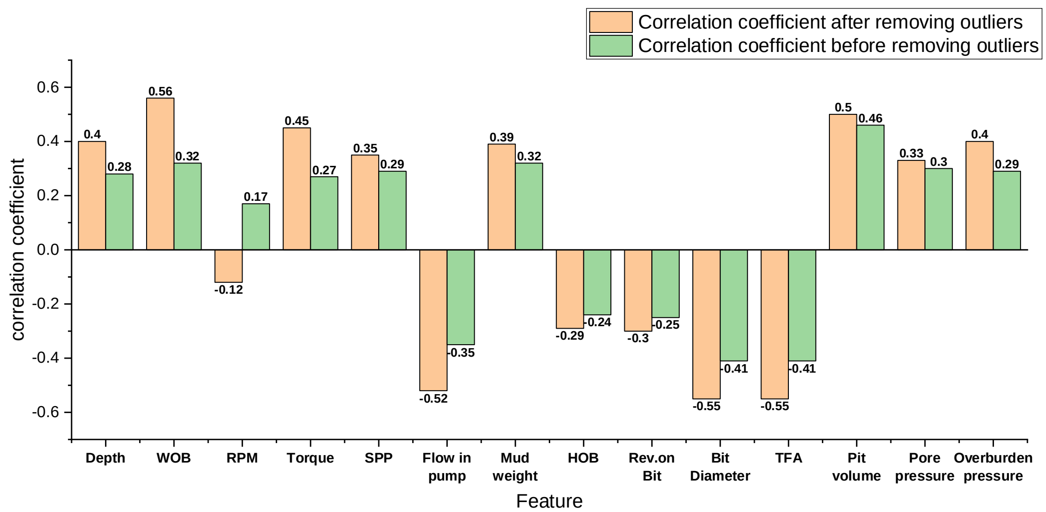

Figure 9. Correlation coefficients for features against the target before and after removing outliers.

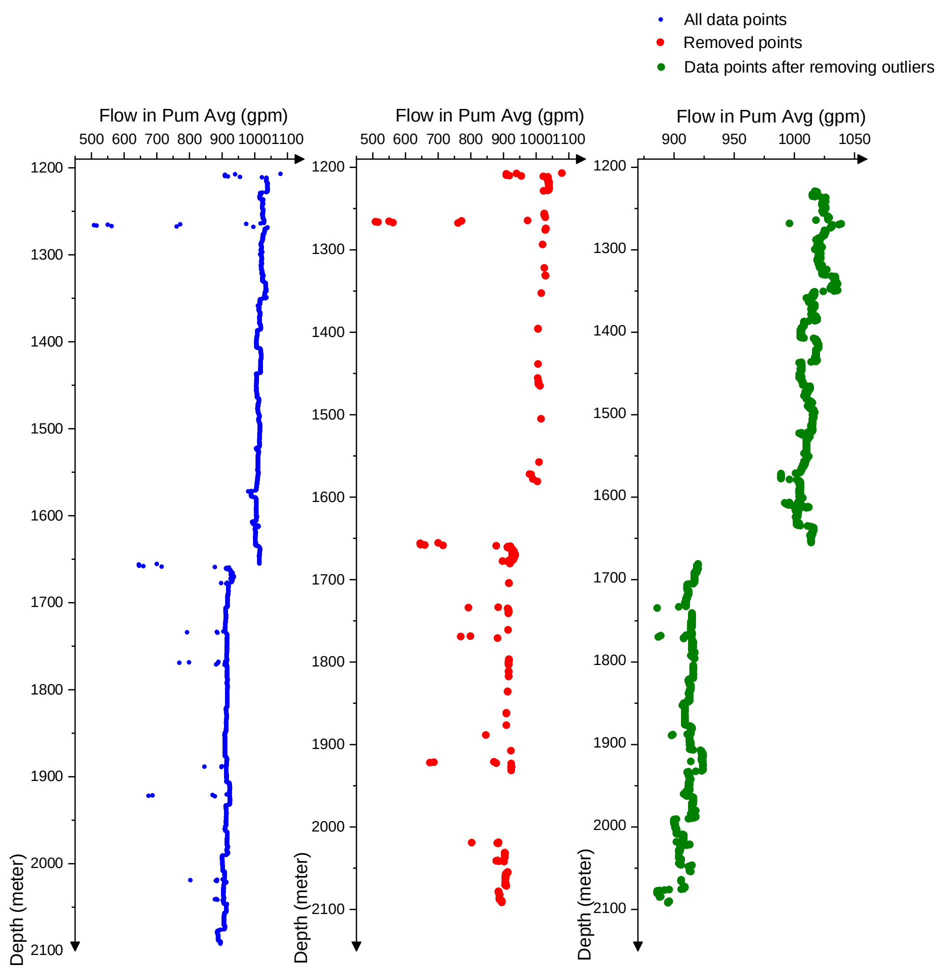

Figure 10. Comparison of average flow in (Q) data points at different preprocessing stages for the dataset.

A training set, a test set, and a validation set are created from the data. The model is trained with 1091 points, tested with 227 points, and validated with 193 points using percentages of 72%, 15%, and 13%, of total data, respectively. The features and the target, as shown in Figure 13, are identified in the model, then the model is built.

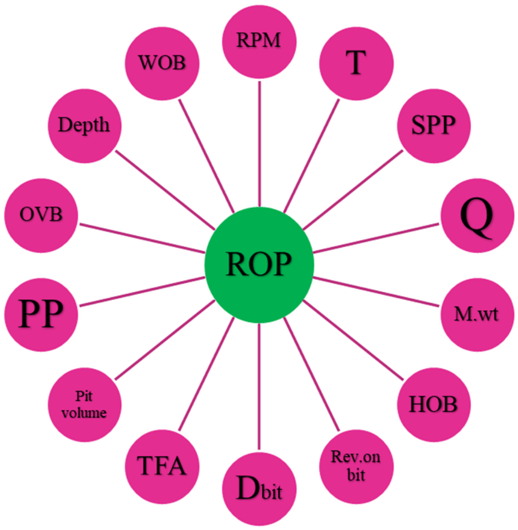

Figure 11. Features included that affect ROP.

3.3. Building model

Two different models were built using two different machine-learning algorithms (KNN and MLP). The models were developed using PYTHON.

3.3.1. KNN model

There are a lot of parameters in KNN the most important two are the number of nearest neighbors (n_neighnors) and weights n_neighbors is set equal to 3 and the weights parameter is used as 'uniform', which means that all points in each neighborhood are weighted equally.

3.3.2. MLP model

A three-layer neural network is used. The input layer contains 14 neurons which is the number of features and the output layer with only one neuron. The hidden layer contains 100 neurons with the rectified linear unit function (relu) as an activation function. lbfgs solver is used as it works well for small datasets.

4. Results and Discussion

Root-mean-square error (RMSE) and coefficient of determination (R^2), which are defined with Equations 1 and 2, respectively, were used to evaluate the statistical analysis of error for each model [31,32].

- **Root-Mean-Square Error (RMSE):**

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(ROP_{actual} - ROP_{predict}\right)^2}{n}} \quad (1)$$

- **Coefficient of determination (R^2):**

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(ROP_{actual} - ROP_{predict}\right)^2}{\sum_{i=1}^{n}\left(ROP_{actual} - \left(\frac{1}{n}\sum_{i=1}^{n} ROP_{predict}\right)\right)^2} \quad (2)$$

In **Table 3**, the statistical results obtained from the two developed models in this paper for training, testing, validation, and total datasets are presented.

Table 3. Results of developed models.

Model	Dataset	RMSE	R^2
KNN	Train	0.606917	0.94326
	Test	1.837651	0.84054
	Validation	1.654272	0.85156
	Total	0.923941	0.91485
MLP	Train	0.265653	0.97502
	Test	2.528778	0.79381
	Validation	1.962094	0.83643
	Total	0.911712	0.92034

For the smaller values of RMSE, it can be inferred that the desired model has higher accuracy in approximation. Additionally, R2 allows for the calculation of the proportion of model outputs that can be explained by a fitted line to the data points. It is shown that KNN works better for test and validation but for total data MLP model has lower RMSE and higher R2.

Figure 14 and **Figure 15** show regression plot of the actual ROP and predicted ROP for training dataset for KNN and MLP models, respectively.

Figure 16 and **Figure 17** show regression plot of the actual ROP and predicted ROP for test dataset for KNN and MLP models, respectively.

Figure 18 and **Figure 19** show regression plot of the actual ROP and predicted ROP for validation dataset for KNN and MLP models, respectively.

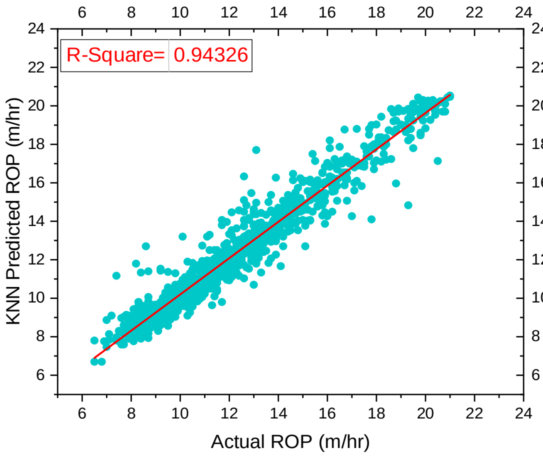

Figure 14. Regression plot of the actual ROP and predicted ROP for training dataset for KNN model.

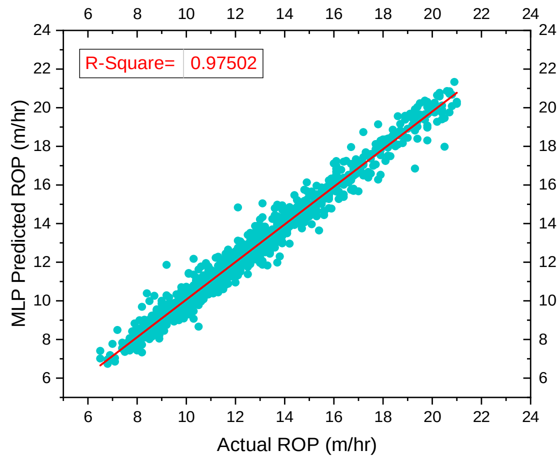

Figure 15. Regression plot of the actual ROP and predicted ROP for training dataset for MLP model.

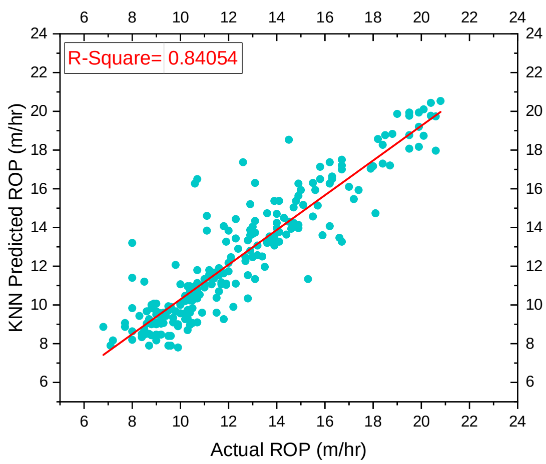

Figure 16. Regression plot of the actual ROP and predicted ROP for test dataset for KNN model.

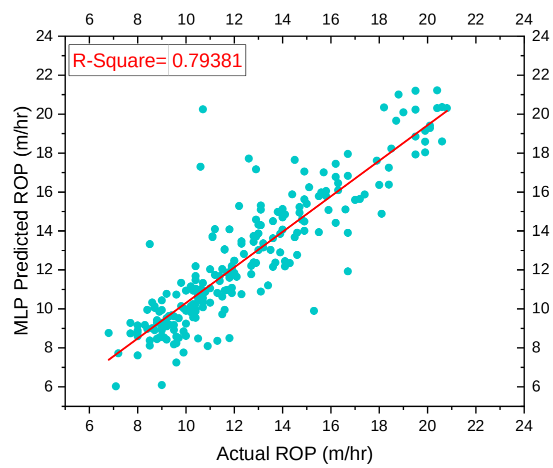

Figure 17. Regression plot of the actual ROP and predicted ROP for test dataset for MLP model.

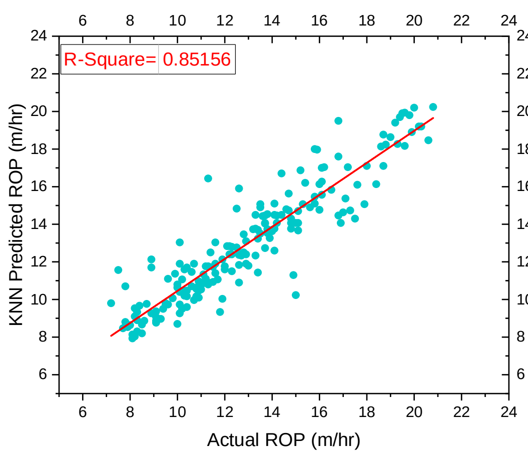

Figure 12. Regression plot of the actual ROP and predicted ROP for validation dataset for KNN model.

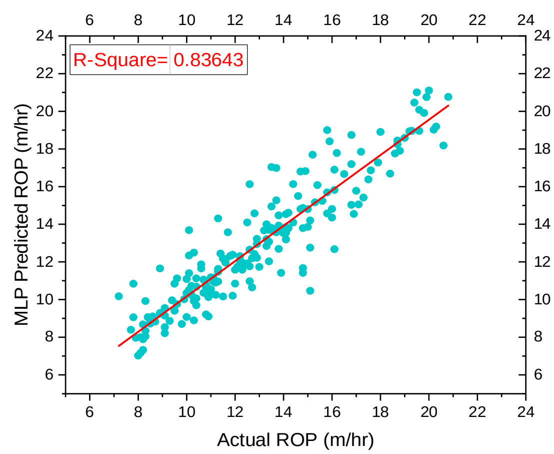

Figure 19. Regression plot of the actual ROP and predicted ROP for validation dataset for MLP model.

According to the results shown above, the MLP model can be considered the best accurate model compared to the other developed model. In addition, it can be deduced from the regression plots that the MLP model has the highest correlation between the predicted and real values.

Figure 20 shows a Comparison between predicted and real ROP among the KNN and MLP models for the total dataset. **Figure 21** and **Figure 22** show a comparison between RMSE and R^2 for both models respectively using a bar chart.

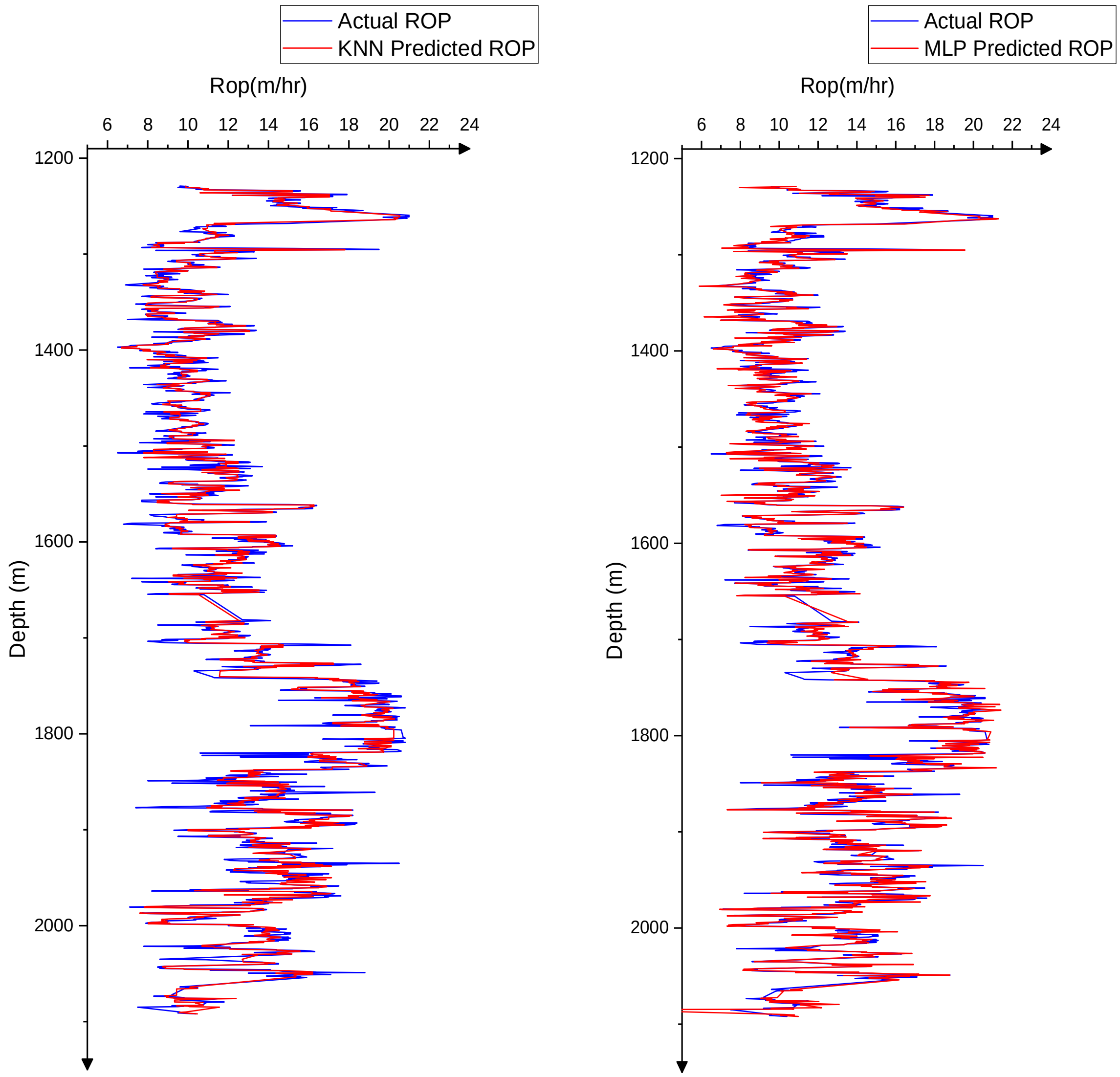

Figure 13. Comparison of predicted and real ROP among the KNN and MLP models for the total dataset.

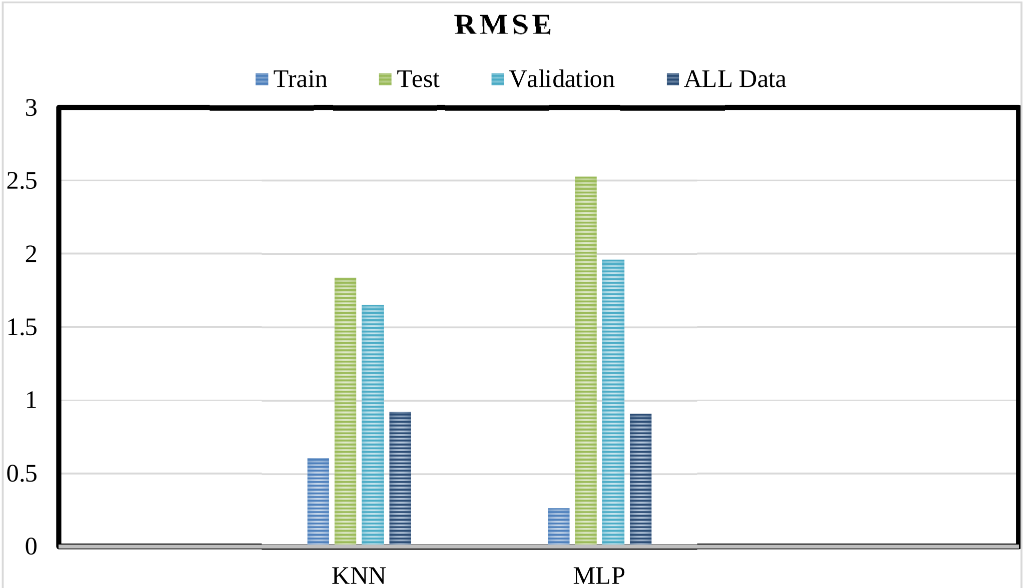

Figure 14. Comparison between RMSE for the two models.

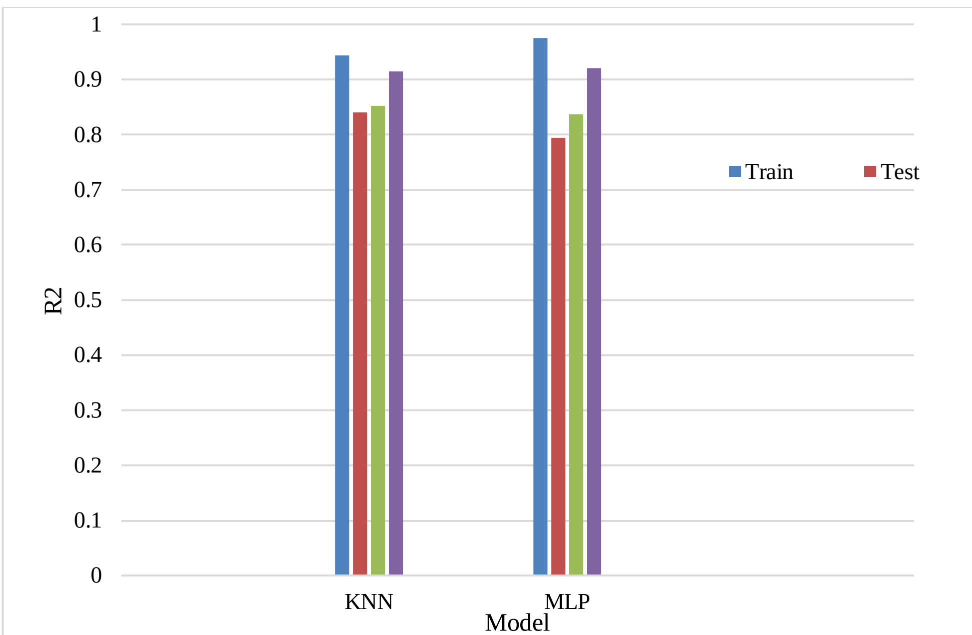

Figure 15. Comparison between R^2 for both models.

5. Conclusions

- Drilling process optimization is closely related to an improvement in the rate of penetration since it can allow faster drilling to a tolerable level while also saving money.
- Process optimization requires knowledge of the connections between the various factors affecting the drilling processes.
- After removing outliers with a percentage of 14.7% of the total data, the correlation coefficients were improved for the features against the target.
- Multi-layer perceptron (MLP) is the best algorithm among the two studied algorithms with accuracy ($R^2 = 92\%$).
- A larger range of data can enable the resulting machine-learning models to have wider applications in certain situations, even though choosing the essential parameters can be difficult.
- Real-time ROP assessment enables the drilling engineer to select the ideal drilling and hydraulic parameters with ease.

- In the case of ROP modeling, machine learning has shown to be significantly more efficient than mathematical models due to the rate at which it can evaluate massive quantities of data across a wide range and the higher accuracy.
- When collecting, organizing, and modeling data, caution must be used. The computer can "learn" wrongly by making correlations between parameters that aren't present because it doesn't understand drilling mechanics.
- The recommended methodology must then be evaluated using data from different oil and gas fields' formations, where drilling operations involve a variety of bit designs, operational considerations, and geological characteristics.

List of abbreviations

ROP	Rate of penetration	MAPE	**Mean absolute percentage error**
ML	Machine learning	RMSE	Root mean squared error
WOB	Weight on bit	ECD	Equivalent circulating density
RPM	Revolution per minute	BPL	Bit pressure loss
T	Torque	D	Well depth
SPP	Standpipe pressure	BRS	Bit rotational speed
Q	Pumping rate	PP	Pump pressure
HOB	Hours on bit	BFR	Bit flow rate
TFA	Total flow area	M.wt	Mud weight
KNN	K-nearest neighbors	FV	Funnel viscosity
$\mathbf{R^2}$	Correlation coefficient	PV	Plastic viscosity
MLP	Multi-layer perceptron	D_{bit}	Bit diameter
BYM	Bourgoyne and Young's model	PP	Pore pressure
ANN	Artificial neural networks	OVB	Over-burden pressure
ReLU	Rectified linear unit	IADC	International association of drilling contractors

Authors' Contributions

Yahia Khamis, methodology, software preparation, validation of software, investigation, and original draft writing. MA, methodology, investigation, and writing review and editing. MN, methodology, investigation, and draft writing. The authors read and approved the final manuscript.

Acknowledgments

All gratitude to Suez University, the Faculty of Petroleum and Mining Engineering, and the staff of Petroleum Engineering department for supporting this research and providing an excellent research environment.

References

[1]. Ahmed, A., Ali, A., Elkatatny, S., and Abdulraheem, A. (2019). New artificial neural networks model for predicting rate of penetration in deep shale formation. Sustainability (Switzerland). 11 (22): 6527.

[2]. Salmani, B., Alanazi, B., Al-Askari, M., Masihi, M., and Paiaman, A. (2009). Effect of Drilling Fluid Properties on Rate of Penetration. Nafta Scientific Journal. 60 (3): 129–134.

[3]. Speer, J.W. (1959). A Method for Determining Optimum Drilling Techniques. Gulf Coast Drilling and Production Meeting.

[4]. Cunningham, R.A. (1960). Laboratory Studies of the Effect of Rotary Speed on Rock-bit Performance and Drilling Cost. Drilling and Production Practice.

[5]. Bingham, M.G. (1964). A New Approach to Interpreting Rock Drillability. Petroleum Publishing Company Co.

[6]. Bourgoyne, A.T. and Young, F.S. (1974). A Multiple Regression Approach to Optimal Drilling and Abnormal Pressure Detection. Society of Petroleum Engineers Journal. 14 (4): 371–384.

[7]. Hareland, G. and Rampersad, P.R. (1994). Drag - Bit Model Including Wear. SPE Journal. 657–667.

[8]. Bourgoyne, A.T., Millheim, K.K., Chenevert, M.E., and Young, F.S. (1986). Applied Drilling Engineering. SPE Textbook, 250 P .

[9]. Anemangely, M., Ramezanzadeh, A., Tokhmechi, B., Molaghab, A., and Mohammadian, A. (2018). Drilling rate prediction from petrophysical logs and mud logging data using an optimized multilayer perceptron neural network. Journal of Geophysics and Engineering. 15 (4): 1146–1159.

[10]. Rabia, H. (2002). Well Engineering &

Construction Hussain Rabia, 789 P .

[11]. Ataei, M., KaKaie, R., Ghavidel, M., and Saeidi, O. (2015). Drilling rate prediction of an open pit mine using the rock mass drillability index. International Journal of Rock Mechanics and Mining Sciences. 73 130–138.

[12]. Farahat, M.S. (2001). Oil Well Drilling Engineering 2, 200 P .

[13]. Nascimento, A., Tamas Kutas, D., Elmgerbi, A., Thonhauser, G., and Hugo Mathias, M. (2015). Mathematical Modeling Applied to Drilling Engineering: An Application of Bourgoyne and Young ROP Model to a Presalt Case Study. Mathematical Problems in Engineering. 2015.

[14]. Hoseinie, S.H., Ataei, M., and Aghababaie, A. (2014). A laboratory study of rock properties affecting the penetration rate of pneumatic top hammer drills. Journal of Mining & Environment. 5 (1): 25–34.

[15]. Dipl.-Ing and Prassl, W.F. (2014). Drilling Engineering, Curtin University of Technology Department of Petroleum Engineering Master, 274 P.

[16]. Hughes, G.R., Hall, W., Pesenti, J., and Ward, P. (2017). Machine learning : the power and promise of computers that learn by example, 125 P .

[17]. Sah, S. (2020). Machine Learning : A Review of Learning Types.

[18]. Habibi Aghdam, H., Jahani Heravi, E., and AG, S.I.P. (2018). Guide to Convolutional Neural Networks A Practical Application to Traffic-Sign Detection and Classification, Springer, 282 P .

[19]. Swamynathan, M. (2017). Mastering Machine Learning with Python in Six Steps, 358 P.

[20]. Burkov Andriy .(2018). The hundred pages Machine learning. Handbook of Research on Cloud and Fog Computing Infrastructures for Data Science, 152 P.

[21]. Pandey, Y.N., Rastogi, A., Kainkaryam, S., Bhattacharya, S., and Saputelli, L. (2020). Machine Learning in the Oil and Gas Industry, 300 P .

[22]. Ali, J.K. (1994). Neural networks: a new tool for the petroleum industry, Proc, Eur. Pet. Comput. Conf. Soc. Pet. Eng.

[23]. Wu, W., Dandy, G.C., and Maier, H.R. (2014). Protocol for developing ANN models and its application to the assessment of the quality of the ANN model development process in drinking water quality modelling. Environmental Modelling and Software. (54): 108–127.

[24]. Ashrafi, S.B., Anemangely, M., Sabah, M., and Ameri, M.J. (2019). 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. 175 604–623.

[25]. Hazbeh, O., Aghdam, S.K. ye, Ghorbani, H., Mohamadian, N., Ahmadi Alvar, M., and Moghadasi, J. (2021). Comparison of accuracy and computational performance between the machine learning algorithms for rate of penetration in directional drilling well. Petroleum Research. 6 (3): 271–282.

[26]. Anemangely, M., Ramezanzadeh, A., Tokhmechi, B., Molaghab, A., and Mohammadian, A. (2020). Drilling rate prediction from petrophysical logs and mud logging data using optimized MLP neural network. J. Phys. Energy. 2 (1): 0–31.

[27]. Sabah, M., Talebkeikhah, M., Wood, D.A., Khosravanian, R., Anemangely, M., and Younesi, A. (2019). A machine learning approach to predict drilling rate using petrophysical and mud logging data. Earth Science Informatics. 12 (3): 319–339.

[28]. Moran, D., Ibrahim, H., Purwanto, A., and Osmond, J. (2010). Sophisticated ROP prediction technologies based on neural network delivers accurate drill time results, Proc, IADC/SPE Asia Pacific Drill. Technol. Conf. Exhib. Held Ho Chi Minh City, Vietnamm, 1–3 November, 100–108.

[29]. Elkatatny, S. (2018). New Approach to Optimize the Rate of Penetration Using Artificial Neural Network. Arabian Journal for Science and Engineering. 43 (11), 6297–6304.

[30]. Hawkins, D.M. (1980). Identification of Outliers. 1st ed. Springer Dordrecht, 188 P.

[31]. Brenjkar, E., Biniaz Delijani, E., and Karroubi, K. (2021). Prediction of penetration rate in drilling operations: a comparative study of three neural network forecast methods. Journal of Petroleum Exploration and Production. 11 (2), 805–818.

[32]. Zhao, Y., Noorbakhsh, A., Koopialipoor, M., Azizi, A., and Tahir, M.M. (2020). A new methodology for optimization and prediction of rate of penetration during drilling operations. Engineering with Computers. 36 (2), 587–595.

پیش‌بینی میزان نفوذ در عملیات حفاری در چاه‌های نفت و گاز توسط K-نزدیک‌ترین همسایه‌ها و الگوریتم‌های پرسپترون چند لایه

یحیی السید*، عادل سالم و شادی الرماح

گروه مهندسی نفت، دانشکده مهندسی نفت و معدن، دانشگاه سوئز، سوئز، مصر

ارسال 2023/02/06، پذیرش 2023/06/01

* نویسنده مسئول مکاتبات: Yahia.Khamis@pme.suezuni.edu.eg

چکیده:

نرخ نفوذ نقش کلیدی در به حداکثر رساندن راندمان حفاری دارد، بنابراین برای بهینه‌سازی و مدیریت فرآیند حفاری ضروری است. مدل‌های ریاضی سنتی با موفقیت برای پیش‌بینی میزان نفوذ در حفاری مورد استفاده قرار گرفته‌اند. به دلیل پیچیدگی زیاد و رفتار غیرخطی پارامترهای حفاری با میزان نفوذ، این مدل‌های ریاضی نمی‌توانند به طور دقیق و جامع میزان نفوذ را پیش بینی کنند. به نظر می‌رسد یادگیری ماشین (ML) جایگزین جذابی برای مدل سازی این فرآیند فیزیکی پیچیده باشد. این مقاله تحقیقاتی مدل‌های جدید مبتنی بر داده را معرفی می‌کند که برای پیش‌بینی ROP با استفاده از پارامترهای مختلف مانند (عمق، وزن روی بیت (WOB)، دور در دقیقه (RPM)، گشتاور (T)، فشار ایستاده (SPP)، جریان در پمپ (نرخ جریان پمپاژ (Q)، وزن گل، ساعت در بیت (HOB)، دور بر روی بیت، قطر بیت، سطح جریان کل (TFA)، فشار منفذ، فشار روباره، و حجم گودال در نظر گرفته شده است. مدل‌های مبتنی بر داده با استفاده از دو ساخته می‌شوند. تکنیک‌های مختلف یادگیری ماشین، با استفاده از 1771 داده میدان واقعی خام و کدگذاری با استفاده از زبان برنامه نویسی پایتون ساخته شده است. مدل k-نزدیکترین همسایه (KNN) که ROP را برای مجموعه داده آموزشی پیش بینی می‌کند ضریب همبستگی 0/94 را نشان می‌دهد. مدل پرسپترون لایه (MLP) که ROP را برای مجموعه داده آموزشی پیش‌بینی می‌کند، ضریب همبستگی 0/98 را نشان می‌دهد. ما می‌توانیم نتیجه بگیریم که MLP دقت بهتری دارد و حذف موارد پرت عملکرد مدل را افزایش می‌دهد.

کلمات کلیدی: نرخ نفوذ، یادگیری ماشین، حفاری، پرسپترون چند لایه، K-نزدیکترین همسایه.