نشریه ژئومکانیک و ژئوانرژی (ژئومکانیک نفت)
GEOMECHANICS and GEOENERGY JOURNAL (GGJ)

Original article

Development of a Machine Learning -Based Framework for Estimating Rate of Penetration in a Hydrocarbon Reservoir

Esmaeil Ghasemi Alamooti[1], Andisheh Alimoradi[2*], Saeedeh Senemari[2], Hassan Reza Ghasemi Tabar[3], Sajjad Talesh Hosseini[2]

1-MSc Student, Department of Mining and Petroleum Engineering, Imam Khomeini International University, Iran

2-Faculty Member, Department of Mining and Petroleum Engineering, Imam Khomeini International University, Iran

3-PhD Candidate, Department of Mining, Petroleum and Geophysics, Shahrood University of Technology, Iran

Received: 2025/05/11; Accepted: 2026/02/07

DOI: 10.22107/ggj.2025.522778.1252

Keywords

Rate of Penetration, Machine Learning, Predictive Modeling, Petroleum Drilling, Drilling Optimization.

Abstract

The primary objective of this research is to enhance the calculation of the Rate of Penetration (ROP) in one of Iran's oil fields using Machine Learning (ML) techniques. Given the significance of ROP evaluation in optimizing drilling operations and reducing costs, this study aims to develop an accurate and efficient model for predicting ROP based on geological characteristics and drilling parameters. AI methods are employed to improve prediction accuracy and optimize drilling processes and ultimately reduce the time and costs of drilling operations. The data required for model generation were collected from one of the Iranian oil fields, encompassing lithological and formation properties and drilling parameters. The back-propagation multi-layer deep models trained, tested and validated with real data. The results demonstrated that ML techniques can significantly enhance ROP prediction accuracy leading to improved and more controlled drilling processes. Additionally, these models can reduce drilling time and substantially lower operational costs. Two ML methods, Random Forest (RF) and Support Vector Machines (SVM), were utilized for modeling based on available data. Input parameters for the model include Depth In-Out/Meterage (m), Weight on Bit (WOB) Min-Max, Rotations Per Minute (RPM) Min-Max, Gallons Per Minute (GPM) Min-Max, Pump Pressure On-Bottom Min-Max, Mud Weight In Min-Max and Mud Weight Out Min-Max. The model's output is the calculated ROP with minimal error.

1. Introduction

Drilling is a critical and costly process in the oil and gas industry. Wells can be categorized into two types based on their purpose: exploratory wells and development (or production) wells. While the former aim to discover hydrocarbons and gather data for development, the latter focus on production. Pressure is typically uncertain during exploratory well drilling, leading to conservative mud weight and casing design practices, which result in low ROP and make optimization challenging.

ROP optimization generally depends on dynamic and static drilling parameters. Dynamic parameters can be controllable for example, Weight on Bit (WOB), Rotations Per Minute (RPM), Mud Flow Rate (Q) or uncontrollable,

* **Corresponding Author:** alimoradi@eng.ikiu.ac.ir.

depending on whether they can be manually adjusted during operations. Static parameters include formation properties such as compressive strength and formation pressure, as uncontrollable parameters. Low ROP prolongs drilling operations and increases costs, while exceeding the optimal ROP may lead to undesirable consequences such as drilling vibrations, rapid bit heating, and dulling. In recent decades, optimizing drilling operations has been a top priority for drilling companies. According to the study (Hassani et al., 2011), alternative models based on advanced mathematical functions combined with AI-based methods can improve the efficiency of horizontal wells in reservoirs by reducing computational costs and increasing production.

In the industry, two approaches exist for predicting ROP in a specific field: physics-based and data-driven. Physics-based models rely on formulas and mathematical functions. Now more than ever, optimizing and achieving the optimal ROP is critical. While several techniques exist for this purpose, each has its advantages and limitations, and no universal model is applicable to all conditions due to the complex nature of parameter relationships. Conventional methods often fail to predict ROP accurately due to the additional complexities of well conditions.

ROP optimization typically involves adjusting WOB, Q, and RPM for efficient drilling. However, ROP has a complex relationship with several other parameters, including formation properties, compressive strength, pressure gradient, mud properties, mud hydraulics, well deviation, and bit size and type. In several cases, increasing WOB and RPM leads to a decrease in ROP due to interactions with formation properties and flow, highlighting the complex underlying relationships between these parameters. To address these uncertainties, data-driven analysis can be effectively used to better understand ROP optimization.

Data-driven techniques have developed greatly in geo and earth science modeling in recent years (Alali et al., 2020; Alimoradi et al., 2011). At the first stages, Artificial Neural Networks (ANNs) were used to predict and estimate geological parameters with simple algorithms (Alimoradi et al, 2020). Later, with the introduction of Machine Learning (ML) and Deep Learning (DL) methods, they were replaced simple ANNs and became more widespread. Algorithms such as Support Vector Machine (SVM), Artificial Bee Colony (ABC) and Random Forest (RF) are among those (Alimoradi et al, 2013; Zhou et al, 2016). In oil industry and especially in drilling rate of penetration, we can see the increasing daily use of artificial intelligence-based methods. Hegde and colleagues predicted the rate of penetration using physics-based and data-driven models (Hegde et al, 2017). They also used machine learning algorithm to perform fully coupled end-to-end drilling optimization (Hegde et al, 2019). Some used sensitivity analysis to detect the most relative parameters to ROP evaluation. He introduced two different parameters as fixed and variable parameters (Barros, 2015; Zhang et al, 2015). A group of scientists used a mathematical modeling to predict performance and wear of PDC drill bits (Mazen et al, 2019; Ahmed et al, 2019). Some others maximized drilling performance with real-time surveillance system and based on parameters optimization algorithm (Meng et al, 2015; Mantha & Samuel, 2016; Saadatmand et al, 2024). Moraveji and Naderi applied optimization algorithms to specifically predict drilling rate of penetration (Moraveji & Naderi, 2016). Others introduced an approach for real-time analytical and machine learning prediction of ROP (Shi et al, 2016; Soares, 2019). Recently, we see the application of hybrid intelligent models which can optimize both input and output data in ROP estimation procedure (Ashrafi et al, 2019; Brenjkar & Biniaz Delijani, 2021). Al-AbdulJabbar and his colleagues used intelligent modeling for ROP estimation in carbonate fractured reservoirs (Al-AbdulJabbar et al, 2020). Tookallo and Sohbatzadeh compared analytical and intelligent methods in ROP estimation in one of the Iranian carbonate reservoirs (Tookallo & Sohbatzadeh, 2023).

The purpose of this research study is to estimate the rate of penetration in petroleum wells using geological and operational attributes. According to the limitation in number of layers, as well as the activation function limit and algorithm processing time in regular neural networks, it is necessary to use the other algorithms with high capabilities such as fast processing time, various activation functions, low layer or single-layer and high accuracy. The reason for introduction and the inherent nature of ML algorithms is to overcome these complexities in such problems that cannot be solved with simple algorithms (Huang et al, 2006). Haj Karimian and his colleagues suggested Random Forest (RF) and Support vector machine (SVM) to overcome these weaknesses (Haj Karimian et al, 2022). Regarding the advantages of these algorithms the ROP value estimation of this research work has been investigated using RF and SVM algorithms.

2. Materials and Method

Studied area is located in the south western part of Iran as an offshore field. According to the limitations, it is impossible to show the detailed data of the petroleum reservoir, but we used the data from one exploration well in this area. All data were extracted from historical drilling information and the necessary matrix for the intelligent matrix were prepared. Total number of 231 data finalized for the model matrix. Input data consist of Depth In-Out, Weight on Bit (WOB), Rotations Per Minute (RPM), gallons Per Minute (GPM), Pump Pressure On-Bottom, Mud Weight In and Mud Weight Out. The output of the model which is our target to be estimated is the bit rate of penetration. Table 1. Shows the part of the data matrix prepared from the historical data of the area.

Table 1. Data Matrix of ML Model

Bit Size	Depth	Meterage	WOB Min-Max	RPM Min-Max	GPM Min-Max	Pump Press on Bottom Min-Max	Mud Weight In Min-Max	ROP
36	13	23	1	30	200	200	65	2.88
36	35	23	6	80	40	40	71	2.88
26	36	120.5	2	50	200	30	65	6.02
26	156.5	120.5	10	130	400	250	71	6.02
171/2	156.5	529.5	0	30	250	130	80	8.15
171/2	686	529.5	25	130	650	2150	96	8.15
121/4	684	0	5000	50	300	650	90	
121/4	684	0	25000	70	315	750	90	
121/4	686.5	1	2000	50	250	780	90	2
121/4	687.5	1	10000	60	270	820	90	2
121/4	687.5	213.5	5	50	60	0	90	7
121/4	710	213.5	15000	120	540	2200	90	7
121/4	687.5	213.5	5	50	60	0	90	7
121/4	901	213.5	15000	120	540	2200	90	7
121/4	901	410	10	90	60	0	90	5.58
121/4	1113	410	25	120	460	2100	93	5.58
121/4	1113	85	5	110	300	700	91	2.39
121/4	1396	85	25	120	460	1350	93	2.39
121/4	1396	330	5	50	420	1270	91	2.67
121/4	1729	330	25	120	500	1750	91	2.67
121/4	1729	247	5	30	400	1200	91	2.53
121/4	1973	247	30	130	500	1860	93	2.53
121/4	1973	0	0	70	400	1240	93	
121/4	1973	0	20	120	500	1900	93	
121/4	1973	0					93	
121/4	1973	0					93	
121/4	1973	0		70	200		93	
121/4	1973	0		70	500		93	
121/4	1973	0	5	30	250	500	93	

2.1. Machine Learning Algorithms

In this study, we utilized two Machine Learning algorithms—Random Forest (RF), and Support Vector Machine (SVM)—to analyze Rate of Penetration of bits (ROP) data. These algorithms were chosen for their varied approaches and demonstrated effectiveness in handling the complexity and high dimensionality typical of drilling datasets.

Random Forest (RF) is powerful ensemble-based method capable of capturing nonlinear interactions among features. RF, by generating numerous decision trees and combining their outputs, offers strong resistance to overfitting and better generalization to new data.

The Random Forest model was implemented with n_estimators=200 to balance computational efficiency and predictive performance. The selected value ensures robust feature interaction capture through aggregated predictions from 200 decorrelated decision trees, while avoiding unnecessary computational overhead. Hyperparameter tuning via out-of-bag error validation confirmed this configuration's stability across multiple bootstrap samples, aligning with the model's inherent resistance to overfitting.

Support Vector Machine (SVM) excels in classification tasks by identifying the optimal hyperplane that separates different classes, even in high-dimensional feature spaces. Its use of kernel functions enables it to handle complex, nonlinear boundaries effectively.

The SVM algorithm was employed with a polynomial kernel to effectively capture nonlinear relationships in the data. This kernel choice enables the model to construct complex decision boundaries by computing feature interactions up to the specified polynomial order.

Through a comparative analysis of these algorithms, we aim to highlight their individual strengths and limitations in the context of ROP data interpretation. This evaluation not only deepens our insight into the influence of different input features on model performance but also guides the selection of the most suitable algorithm for specific drilling conditions. Ultimately, our work seeks to support more accurate subsurface characterizations, thereby enhancing decision-making in petroleum drilling.

2.2. Random Forest

Random Forest, also known as Random Decision Forest, is an ensemble learning method used for both classification and regression tasks. It operates by constructing a multitude of decision trees during training and outputs either the most frequent class (in classification) or the average prediction (in regression) from all trees (Fukunaga and Hostetler, 1975). This method is particularly effective when individual decision trees are pre-fitted on training subsets.

One of the key advantages of Random Forest is its simplicity and minimal tuning requirements—it primarily depends on two parameters: the number of trees in the forest and the number of features considered at each split. Notably, the algorithm demonstrates robustness, showing relatively low sensitivity to these parameters (Yizong, 1995).

As illustrated in Fig.1, Random Forest leverages the collective decision-making of multiple trees to enhance performance. By averaging the outputs of individual trees, it reduces variance, mitigates overfitting, and improves predictive accuracy (Dorin and Meer, 2002)

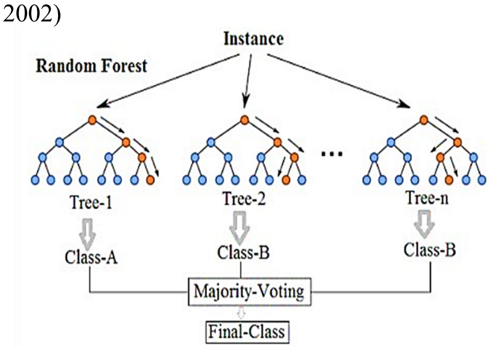

Fig.1. Random forest performance procedure (Simorgh et al, 2024)

2.3. Support Vector Machine

Support Vector Machine (SVM) is a supervised learning algorithm renowned for its strong generalization capabilities across diverse problem domains. It is well-suited for both binary classification and regression tasks. The algorithm functions by constructing a hyperplane in an n-dimensional space that effectively separates different data classes.

This hyperplane is optimized to maximize the margin—the distance between the closest data points of each class—which enhances the model's ability to generalize (Ashok Kumar, 2020). In essence, SVM seeks the boundary that optimal separates the categories while maintaining the greatest possible distance from all data points.

Fig.2 demonstrates how the SVM algorithm distinguishes between groups of similar data using this margin-maximizing approach.

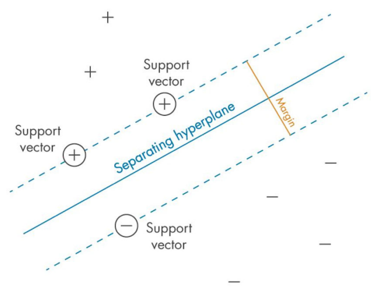

Fig.2. Support vector machine performance (Liu et al, 2009)

3. Results and Discussion

In this study, two different ML techniques were used to predict ROP using the input parameters listed in Table 1, which include WOB, SPP, RPM, GPM, MW_in, T, and GR. The Random Forest and Support Vector Machine algorithms were trained using 75% of the well data, while 25% was used for testing and validation.

The training dataset was randomly selected and then reviewed to cover four different formations penetrated during drilling. Two models were developed by performing sensitivity analyzes for the design parameters of each model, and model performance was evaluated by calculating the Root Mean Square Error (RMSE), Mean Squared Error (MSE) and R^2 score. Two ML tools were used for ROP prediction, both widely applied in AI applications in the oil industry. RF, as a tool, has key controlling parameters such as the number of layers, number of neurons in hidden layers, training function, and transfer function. Optimizing these parameters enhances model accuracy. SVM was used as a tool for predictive models.

Modeling was conducted based on the two aforementioned techniques. Fig.3 illustrates the preprocessing and initial modeling of the data.

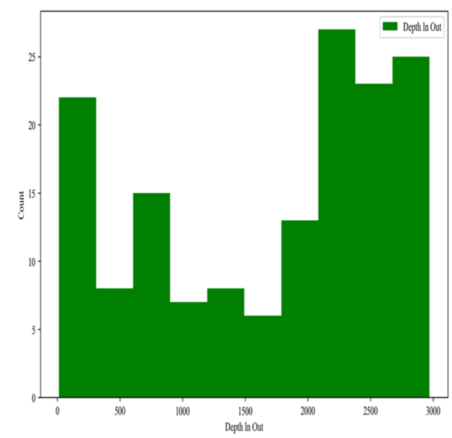

Fig.3. Preprocessing and Initial Data Modeling

This histogram displays the distribution of the number of cases (Count) relative to Depth In Out. The chart addresses the number of cases at each output depth. Key observations and overall analysis include:

- Data Distribution: The highest frequency corresponds to higher depths (approximately 2500–3000), indicating that most data points are in this range, suggesting a focus on deeper depths. Conversely, the number of cases at shallow depths (below 500) is significantly higher, with peaks indicating data concentration at shallow depths.

- Trends and Patterns: A fluctuating pattern in the number of cases is observed, with a high count initially (0–500), followed by a decrease and subsequent increase. This fluctuation may reflect factors related to data characteristics or specific trends in the studied domain.

- First Peak: At the start of the x-axis (0–500), the number of cases is very high, indicating a significant portion of data at shallow depths.

- Gradual Decrease: After this peak, the number of cases gradually decreases until around 1500, suggesting reduced data distribution at intermediate depths.

- Second Peak: An increase is observed at higher depths, particularly around 2500, creating another peak. These oscillatory patterns may indicate structural or procedural differences in the data.

This unstable distribution may require further investigation to understand technical or operational reasons. If related to comparing depths in sample data, the results may indicate where most cases are located. In cases focusing on shallow depths (for example, drilling or surface layer vulnerability studies), the results highlight the dominance of surface nodes. To improve modeling or analysis, it is necessary to determine why data are dispersed across a wide range of depths and, if needed, adjust sampling or data correction.

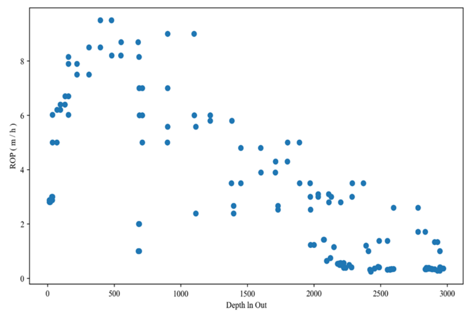

Fig.4. Preprocessing and Initial Data Modeling – ROP vs. Depth

The plot in Fig.4 illustrates the relationship between Depth In Out and ROP (m/h). A detailed and scientific analysis follows:

- Inverse Relationship: Generally, an increase in depth is associated with a decrease in ROP. Data points are concentrated at higher ROP values at shallow depths but shift to lower ROP values as depth increases.

- Trends and Patterns: At shallow depths (0–500), ROP values typically range from 5 to 9 m/h, with a high number of points. At intermediate to high depths (>1500), scatter decreases, and points cluster at lower ROP values, especially below 2 m/h. At depths above 2500, most points are at ROP values below 2, indicating reduced activity or efficiency at greater depths.

- Correlation: A significant inverse relationship exists between depth and ROP, where ROP tends to decrease as depth increases. This is common in drilling, mining, or fluid transfer projects, where efficiency or speed decreases at greater depths.

- Technical Implications: The decrease in ROP with depth may indicate technical or physical limitations of equipment, rock resistance, or reduced productivity at greater depths. Improving performance may require advanced drilling technologies or specific maintenance measures at deeper depths.

Fig.5 illustrate the implementation of Random Forest algorithm on train data.

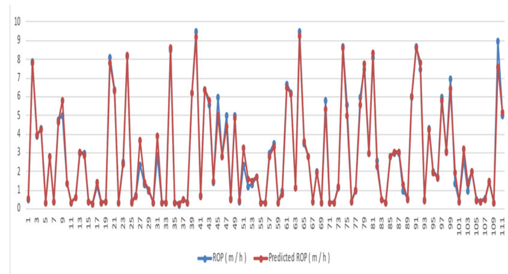

Fig.5. RF Model for Actual and Predicted ROP in Train Data

As it can be seen from this figure, the predicted ROP values follow the actual data very closely. The error value of estimation is very low and the RMSE is about 0.4 which indicates that the model has learned the training patterns extremely well. Such a near-perfect fit can be attributed to RF's ability to model complex nonlinearities, but it also raises the potential for overfitting which should be checked during test and validation process.

For Support Vector Machine, the results of applying the model on training data is shown in Fig.6. The algorithm, exactly applied for the same training data set as RF model in Fig.5.

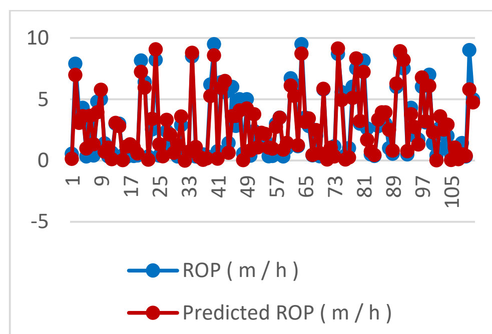

Fig.6.SVM Model for Actual and Predicted ROP in Train Data

Results show that the SVM-predicted values are generally aligned with the actual ROP values but show slightly more deviation compared to RF. This implies a moderately higher RMSE than RF. The RMSE value is 0.7 which indicates the better training process for RF than SVM. The SVM model does not overfit the training data, likely due to margin maximization, which inherently guards againstoverfitting. This could

limitations in capturing more nuanced variations. At the next stage, the models should be applied for test data set to check whether RF or

SVM performs better. Fig.7 shows the application of RF model on test data.

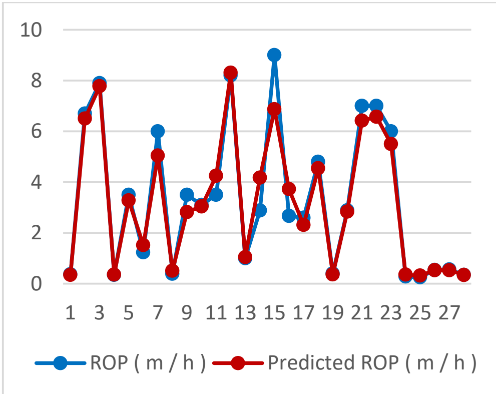

Fig.7. RF Model for Actual and Predicted ROP in Test Data

The algorithm predictions continue to match the actual ROP values closely with small, controlled deviations with RMSE of about 0.5 which indicates that overfitting is minimal and the model generalizes well. The RMSE appears slightly higher than in training—an expected behavior of a well-regularized model. RF is averaging over trees allows it to maintain robustness against noise, which is evident here.

In Fig.8, the results of applying SVM algorithm on test data is shown. It can be seen that SVM's performance is comparable but slightly less accurate than RF on unseen data and some data points show larger deviations, suggesting higher RMSE than RF in testing. The RMSE value is about 0.8 which may stem from kernel choice limitations or insufficient tuning of C and epsilon parameters for the regression margin. Nonetheless, the model shows good generalization, likely due to its inherent structural risk minimization

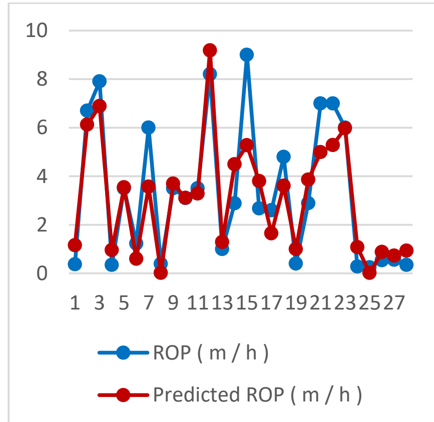

Fig.8. SVM Model for Actual and Predicted ROP in Test Data

Results show that Random Forest is more robust in estimating test data which were unfamiliar for the algorithm. If the network is stable enough, the same result should be obtained for validation data which are completely unfamiliar for the model. Fig.9 and 10 show the results for validation data for RF and SVM accordingly.

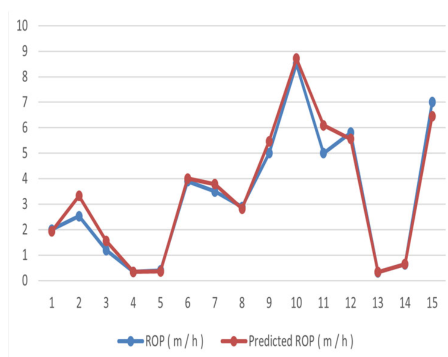

Fig.9. RF Model for Actual and Predicted ROP in Validation Data

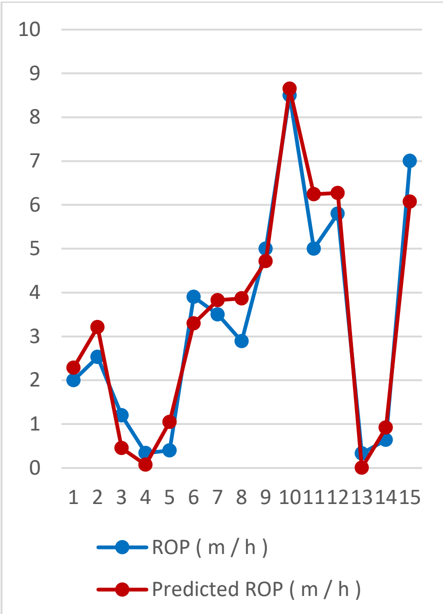

Fig.10. SVM Model for Actual and Predicted ROP in Validation Data

It is very interesting that RF maintains a strong performance, with predicted and actual ROP values still in close alignment. This consistency across all three datasets confirms that RF is well-tuned, with low RMSE (0.6) and good generalization. RF's capacity to capture complex relationships through ensemble learning is evident here.

For SVM, the predicted ROP shows a reasonable trend alignment with actual values but with noticeable scatter at some points. This indicates a modest increase in RMSE (near 1.1), possibly due to data points near the decision boundary or where kernel generalization weakens.

Table 2. shows the RMSE values for each model separately.

Table.2. RMSE,MSE and R^2 score Values for RF and SVM in Training, Testing and Validation Dataset

model	Dataset	Mean squared error	R2 score	RMSE
SVM	Train	0.006894638	0.92114741	0.6-0.8
	Test	0.01650172	0.82095208	0.7-0.9
	valid	0.004623741	0.9362415	0.9-1.2
RandomForest (RF)	Train	0.001119456	0.987197	0.3-0.5
	Test	0.004345918	0.95284567	0.4-0.6
	valid	0.002083773	0.97126607	0.5-0.7

In this study, we developed and evaluated two distinct Machine Learning frameworks—Random Forest (RF) and Support Vector Machine (SVM)—to estimate the Rate of Penetration (ROP) in one of Iran's hydrocarbon reservoirs. The objective was to identify a robust and accurate predictive model that can enhance the interpretation of drilling performance, which is crucial for optimizing operational efficiency and reducing drilling costs.

The models were trained, tested, and validated on real-world drilling datasets. Their performances were evaluated both qualitatively through visual comparison and quantitatively using Root Mean Square Error (RMSE) estimates. The results of six comparative graphs—covering training, testing, and validation datasets—highlighted critical insights into the behavior and suitability of each model in handling used data.

The Random Forest model demonstrated superior performance across all datasets. It exhibited a strong ability to capture complex, nonlinear relationships inherent in drilling parameters such as lithology, weight on bit, mud properties, and rotation speed. With visually tight alignment between predicted and actual ROP values, its RMSE ranged from 0.3–0.5 m/h in training, 0.4–0.6 m/h in testing, and 0.5–0.7 m/h in validation, indicating a high degree of accuracy and generalization. The ensemble nature of RF—aggregating multiple decision trees—enabled it to effectively minimize variance and control overfitting, which proved beneficial in modeling the noisy and heterogeneous nature of subsurface data.

In contrast, the Support Vector Machine model, while effective in principle, showed moderate predictive capability relative to RF. Its RMSE values were 0.6–0.8 m/h for training, 0.7–0.9 m/h for testing, and 0.9–1.2 m/h for validation, reflecting a gradual decline in performance as the model was exposed to new data. This behavior can be attributed to the sensitivity of SVM to kernel selection and hyperparameters, especially in high-dimensional, nonlinearly separable datasets typical of geological environments. Though SVM exhibited aconsistenttrend-followin behavior, the prediction error increased under unseen conditions,indicatingpotential underfitting or ins ufficient kernel flexibility.

The comparative evaluation underscores that Random Forest is more suitable for this ROP prediction task, particularly when the objective is to balance accuracy, generalization, and model interpretability. Its resilience to outliers and noise, coupled with its minimal requirement for hyperparameter fine-tuning, makes it an ideal candidate for deployment in real-time drilling optimization systems.

Future work could involve expanding the dataset with more diverse drilling conditions, exploring more advanced models such as XGBoost, deep neural networks, or LSTM architectures for time-series prediction, and embedding these models into intelligent drilling advisory systems. Additionally, evaluating model interpretability through feature importance analysis or SHAP values would provide deeper insight into the drilling parameters that most influence ROP, which can be highly valuable for field engineers and geoscientists.

4.Aknowledgment

Any help and support through providing the required data for this research is strongly appreciated.

5.References

[1] Ahmed, O. S., Adeniran, A. A., & Samsuri, A. (2019). Computational intelligence-based prediction of drilling rate of penetration: A comparative study. Journal of Petroleum Science and Engineering, 172, 1–12.

[2] Alali, A. M., Abughaban, M. F., & Amanc, B. M. (2020). Hybrid data driven drilling and rate of penetration optimization. Journal of Petroleum Science and Engineering.

[3] Al-AbdulJabbar, A., Elkatatny, S., Mahmoud, A. A., Moussa, T., AlShehri, D., Abughaban, M., & Al-Yami, A. (2020). Prediction of the rate of penetration while drilling horizontal carbonate reservoirs using the self-adaptive artificial neural networks technique.

[4] Alimoradi, A., Angorani, S., Ebrahimzadeh, M., & Shariat Panahi, M. (2011). Magnetic inverse modeling of a dike using the artificial neural network approach. The Journal of Near Surface Geophysics (EAGE), 9, 339–347.

[5] Alimoradi, A., Maleki, B., Karimi, A., Sahafzadeh, M., & Abbasi, S. (2020). Integrating geophysical attributes with the new cuckoo search machine-learning algorithm to estimate silver grade values– Case study: Zarshouran gold mine. Journal of Mining and Environment, 11(3), 865–879.

[6] Alimoradi, A., Moradzadeh, A., & Bakhtiari, M. R. (2013). Application of artificial neural networks and support vector machines for carbonate pores size estimation from 3D seismic data. Journal of Mining and Environment, 4(1), 1–14.

[7] Ashrafi, S. B., Anemangely, M., Sabah, M., & 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.

[8] Barros, L. (2015). ROP modeling chronology, sensitive analyses, and field data comparisons [Master's thesis, The University of Texas at Austin].

[9] Brenjkar, E., & Biniaz Delijani, E. (2021). Prediction of penetration rate in drilling operations: A comparative study of three neural network forecast

methods. Journal of Petroleum Exploration and Production Technology, 11, 805–818.

[10] Davari, M. A., Senemari, S., Alimoradi, A., & Safavi, S. J. (2024). Permeability prediction from log data using machine learning methods. Journal of Petroleum Geomechanics. https://doi.org/10.22107/JPG.2024.426878.1220

[11] Dorin, C., & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(5), 603–619.

[12] Fukunaga, K., & Hostetler, L. D. (1975). The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Transactions on Information Theory, 21(1), 32–40.

[13] Ghasemitabar, H. R., Alimoradi, A., Hemmati Ahooi, H. R., Fathi, M., & Sarookhani, M. (2024). Intelligent borehole simulation with Python programming. Journal of Mining and Environment. https://doi.org/10.22044/JME.2025.15310.2936

[14] Ghasemitabar, H., Rouhani, A. K., Arab-Amiri, A. R., & Moosavi, M. (2026). Using reinforcement learning in mineral exploration: Integrating ASTER, Landsat 9, and geochemical data while comparing machine learning techniques. China Geology, 2024(1),104.https://doi.org/10.31035/cg2025223

[15] Ghasemitabar, H. R., Talesh Hosseini, S., Alimoradi, A., Fathi, M., & Sahafzadeh, M. (2025). Intelligent borehole simulation with Python programming. Journal of Mining and Environment, 15(2). https://doi.org/10.22044/jme.2023.13610.2527

[16] Haj Karimian, H., Alimoradi, A., Hemmati Ahoei, H. R., & Salsabili, M. (2022). Comparison between the performance of four metaheuristic algorithms in training a multilayer perceptron machine for gold grade estimation. International Journal of Mining and Geo-Engineering, 56(2), 97–105.

[17] Hassani, H., Sarkheil, H., Foroud, T., & [et al.]. (2011). A proxy modeling approach to optimize horizontal well placement. 45th US Rock Mechanics/Geomechanics Symposium. OnePetro.

[18] Hegde, C., Daigle, H., Millwater, H., & Gray, K. (2017). Analysis of rate of penetration (ROP) prediction in drilling using physics-based and data-driven models. Journal of Petroleum Science and Engineering, 295–306.

[19] Huang, G., Yu, Z., & Kheong Sew, C. (2006). Extreme learning machine: Theory and applications. Neurocomputing, 70, 489–501.

[20] Liu, H., Wen, S., Li, W., Xu, C., & Hu, C. (2009). Study on identification of oil/gas and water zones in geological logging based on support vector machine. Fuzzy Information and Engineering, 2, AISC 62, 849–857.

[21] Mantha, B., & Samuel, R. (2016). ROP optimization using artificial intelligence techniques with statistical regression coupling. SPE Annual Technical Conference and Exhibition, Dubai, UAE.

[22] Mazen, A. Z., Mujtaba, I. M., Hassanpour, A., & Rahmanian, N. (2019). Mathematical modelling of performance and wear prediction of PDC drill bits: Impact of bit profile, bit hydraulic, and rock strength. Journal of Petroleum Science and Engineering, 106849.

[23] Meng, C., Mengci, S., Jinwen, Z., Kai,

K., & Yucai, L. (2014). Maximizing drilling performance with real-time surveillance system based on parameters optimization algorithm. Advances in Petroleum Exploration and Development, 8, 15–24.

[24] Moraveji, M. K., & Naderi, M. (2016). Drilling rate of penetration prediction and optimization using response surface methodology and bat algorithm. Journal of Natural Gas Science and Engineering, 31, 829–841.

[25] Nazarkhanloo, A., Alimoradi, A., Senemari, S., Ghasemitabar, H., & Talesh Hosseini, S. (2026). Using the You only Look once Framework for Enhanced Plagioclase Classification in Petrographic Thin Sections. Mining, Metallurgy and Exploration. https://doi.org/10.1007/s42461-026-01481-8

[26] Razzaghzadeh, R., Shakoor Shahabi, R., & Nouri Qarahasanlou, A. (2019). Analysis of reliability and maintainability for multiple repairable units (Case study: Sungun copper mine). Journal of Mining and Environment, 10(4), 1045–1059.

[27] Saadatmand, F., Alimoradi, A., Moosavi, M. H., Davari, M. A., & Rezakhani, P. (2024). Optimizing deep learning models for shear wave velocity estimation utilizing petrophysical logs: A case study on an oil reservoir in Southern Iran. Journal of Petroleum Geomechanics. https://doi.org/10.22107/JPG.2024.436270.1225

[28] Senemari, S. (2017). Calcareous nannofossils biostratigraphy of the Gurpi Formation at Kalchenar section in Northwest of Izeh. Journal of Geoscience, 26(103), 287–294.

[29] Shi, X., Liu, G., Gong, X. L., Zhang, J. L., Wang, J., & Zhang, H. N. (2016). An efficient approach for real-time prediction of rate of penetration in offshore drilling. Mathematical Problems in Engineering, 3575380.

[30] Simorgh, M., Alimoradi, A., Hemmati Ahooi, H., Salsabili, M., Fathi, M., Ghasemi Tabar, H., & Rezakhani, P. (2023). Application of supervised machine learning inversion in the estimation of iron ore grade from geophysical data: Comparative investigation of GB, RF and SVM algorithms. Journal of Geomine, 1(3), 123–136.

[31] Soares, C., & Gray, K. (2019). Real-time predictive capabilities of analytical and machine learning rate of penetration (ROP) models. Journal of Petroleum Science and Engineering, 172, 934–959.

[32] Tookallo, K., & Sohbatzadeh, S. (2023). Using analytical and artificial intelligence models to estimate drilling rate of penetration (ROP). Petro Chem Indus Intern, 6(5), 322–327.

[33] Yizong, C. (1995). Mean shift, mode seeking, and clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 17(8), 790–799.

[34] Zarei, M., Shahabi, R. S., Hadei, M. R., et al. (2022). The use of air decking techniques for improving surface mine blasting. Arabian Journal of Geosciences, 15(19), 1–12. https://doi.org/10.1007/S12517-022-10826-8

[35] Zhang, J., Li, Y., & Liu, Z. (2015). Optimization of drilling parameters for rate of penetration improvement in directional wells. Journal of Petroleum Science and Engineering, 133, 202–212.

[36] Zhou, H., Zhang, S., & Li, Z. (2016). Prediction of drilling rate of penetration using a hybrid support vector regression model. *Journal of Natural Gas Science and Engineering, 33*, 1–11.