Shahrood University of Technology

Journal of Mining and Environment (JME)

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

Iranian Society of Mining Engineering (IRSME)

Leveraging Schimazek's Abrasivity Index and Drilling Fluid Conductivity for Enhanced ROP Prediction: A Comparative Machine Learning Study

Shahrokh Khosravimanesh[1], Akbar Esmaeilzadeh[2], Reza Mikaeil[2*], and Sasan Ghorbani[3]

1. Department of Mining Engineering, Isfahan University of Technology (IUT), Isfahan, Iran
2. Department of Mining Engineering, Environment Faculty, Urmia University of Technology (UUT), Urmia, Iran
3. Department of Mining Engineering, Hamedan University of Technology, Hamedan, Iran

Article Info

Received 23 February 2026
Received in Revised form 18 July 2026
Accepted 31 August 2026
Published online 31 August 2026

DOI: 10.22044/jme.2026.17539.3489

Keywords

Rate of penetration
Electrical conductivity
Huber regression
Elastic network
Decision tree

Abstract

The rate of penetration (ROP) is a key indicator in drilling operations, directly influencing cost estimation and project scheduling. Direct measurement of ROP is typically costly and limited in applicability, therefore developing accurate data-driven predictive models is of great importance. In this study, a dataset of 492 laboratory measurements was collected, incorporating four independent variables-Schimazek's Abrasivity Index (SFa), drilling fluid electrical conductivity (EC), weight on bit (WOB), and bit rotation speed (BRS)-while ROP was considered as the dependent variable. Four supervised machine learning algorithms, namely Huber regression, Elastic Net, Ensemble, and Decision Tree, were employed to predict ROP. The predictive performance of the models was evaluated using MAPE, R2, RMSE, and EVS metrics. Additionally, a composite objective function integrating these indices was defined to provide an overall ranking of the models. The results revealed that the Decision Tree model achieved the best performance with an objective function value of 0.2975, followed by the Ensemble, Elastic Net, and Huber regression models with values of 0.5655, 1.4471, and 1.2702, respectively. These findings demonstrate that the Decision Tree algorithm offers the most accurate and reliable predictions of ROP and can serve as an effective tool for optimizing drilling time and reducing operational costs.

1. Introduction

Drilling is one of the most important and expensive processes in mining projects and its success has a direct impact on equipment productivity, project execution time and total operational costs. In this process, the rate of penetration (ROP) is used as a fundamental indicator of efficiency and cost prediction. The decrease of ROP results in the increase of drilling time and the direct and indirect costs of the project. The correct prediction of ROP allows optimal planning, risk reduction and significant savings of financial and time resources. ROP is not only a criterion that affects the speed of drill advancement, but also an important indicator for the assessment of equipment performance, the determination of productivity of human resources and the optimization of drilling process. Due to the mechanical and abrasive complexities of rock, different operating conditions, environmental factors, and also drilling induced vibrations and fluid and rock interactions, managing and predicting ROP is one of the main challenges in the mining and drilling industry, which directly impacts the economic success of mining projects [1–6].

Direct measurement of ROP through field and laboratory experiments, although providing valuable information, is very expensive, time-consuming, and limited to specific conditions. In addition, generalization of the results of experiments to diverse geological environments is associated with uncertainty and error. In such

Corresponding author: reza.mikaeil@gmail.com (R. Mikaeil)

circumstances, the development of data-driven predictive models is proposed as an efficient approach. These models are able to provide a reliable estimate of ROP by utilizing measurable parameters, such as mechanical and abrasion properties of the rock, drilling operating conditions, and drilling fluid properties, and are a valuable alternative to expensive direct methods.

In recent years, numerous analytical, statistical, and numerical models have been developed for predicting ROP. Although these models have improved the understanding of drilling mechanisms, their limitations in representing the complex and nonlinear relationships between drilling parameters and penetration rates have motivated the development of advanced data-driven approaches. The most important achievements in this field are reviewed in the following section.

Mathematical models are among the earliest approaches used to predict the ROP, and are developed based on analytical relationships between drill bit load, rotation speed, and rock properties. By simplifying the rock fracture mechanism, these models establish relationships between the mechanical and abrasive properties of the rock and the penetration rate. Previous studies have also explored the effect of the intrinsic properties of the geological materials on their engineering behavior. Indices such as mechanical specific energy (MSE) and other analytical correlations have revealed the influence of rock properties on drilling efficiency and penetration rate, and optimization techniques have been applied to improve the predictive performance of conventional ROP models by calibrating parameters. The main advantages of these models are their simplicity, rapid applicability and direct interpretation of the relationship between rock properties and penetration rate. But their accuracy is severely reduced under complex geological conditions, in heterogeneous or highly abrasive rocks, and when the effects of drilling fluids and the environment are important. Such limitations led to the development of statistical models based on empirical data. [7–15].

Statistical models have been developed to predict the ROP based on field and laboratory data, with the aim of quantifying the relationship between rock properties and penetration rate. In such models, multivariate regression, correlation analysis and empirical indices (e.g., Rock Drillability Index (RDi) and Drilling Rate Index (DRI)) are commonly used to assess the effect of rock mechanical and abrasion properties, and drilling conditions on ROP. The results have shown the important role of empirical parameters in the ROP model. The advantages of these models include computational transparency, interpretability of the results and the possibility of fast calibration with real data. However, they are very sensitive to the quality of input data and often have low accuracy when applied to new geological conditions or data outside the training range. These shortcomings have encouraged the development of numerical models and, in particular, data-driven algorithms [3, 16–26].

Numerical models have been developed to simulate the mechanisms influencing ROP and predict drilling performance under various conditions. Finite Element Method (FEM), Discrete Element Method (DEM) and other such methods have been extensively used to study the influence of rock geomechanical properties, in-situ stresses, loading conditions and rock failure mechanisms on penetration rate. These models have improved the understanding of drilling behavior by providing detailed information on the mechanisms of rock fragmentation and bit-rock interaction. The main advantage of numerical models is their physical representativeness and the detailed mechanical insight into the drilling process. However, they are limited by the need for accurate and extensive input data, the high computational cost, and the difficulty of extrapolating the results to other geological settings. These restrictions have compelled researchers to use and develop data-driven and machine learning algorithms for ROP prediction [27–37].

In recent years, machine learning (ML) models have been developed as a sophisticated framework for ROP prediction with higher accuracy and generalization ability than mathematical, statistical and numerical models. The models are able to capture the complex non-linear relationships between rock properties, drilling fluid condition and operational parameters and thus enable highly accurate ROP predictions.

Previous studies have shown the promising potential of machine learning techniques for ROP prediction. The data-driven algorithms had a greater ability to detect complex relations among drilling parameters compared to traditional methodologies [38]. Various ML techniques such as artificial neural network (ANN), Adaptive Neuro-Fuzzy Inference Systems (ANFIS), Support Vector Regression (SVR), boosting algorithms, ensemble models and hybrid optimization-based approaches have been successfully used to improve

the prediction accuracy and the generalization of the model under different geological and operational conditions [39–49]. For instance, K-nearest neighbors (KNN) and multi-layer perceptron (MLP) algorithms have been used for ROP prediction for drilling operations, demonstrating the feasibility of data-driven approaches for estimating drilling performance [50]. Intelligent approaches based on experimental datasets have also been used to predict drilling performance parameters in rock drilling operations [51]. In addition, studies based on both laboratory and field data have confirmed the effectiveness of ML models in capturing drilling behavior and providing reliable ROP predictions in practical applications [22,52].

Overall, the existing literature highlights the strong potential of machine learning for ROP modeling and prediction, owing to its ability to capture complex relationships among drilling parameters. Consequently, recent research has increasingly focused on the development and refinement of ML-based approaches for improving prediction performance [23–59].

Including drilling fluid electrical conductivity (EC) as a predictive variable requires a physical explanation. EC is closely related to the ionic concentration of the drilling fluid and may influence ROP through several mechanisms. First, increased ionic concentrations can enhance lubrication at the bit-rock interface, which can decrease friction and decrease the energy required to fracture rock. Secondly, fluids with higher ionic content are typically more effective in heat transfer; this helps to dissipate heat produced by friction and maintain drilling efficiency. Third, EC can be considered an indicator of the chemical activity of the drilling fluid, which may affect rock softening and facilitate crack propagation during drilling. These mechanisms suggest that EC may influence drilling performance and provide a physical basis for its inclusion as an input variable in ROP prediction models. Therefore, EC was considered as a novel parameter in the present study.

Based on previous studies and the critical importance of accurately estimating the rate of penetration (ROP) in rocks, the present study selected rock samples from various regions and compiled a dataset comprising 492 laboratory observations. Four independent variables, including Schimazek abrasivity index (SFa), drilling fluid electrical conductivity, weight on bit (WOB), and bit rotation speed (BRS), were used as model inputs, while ROP was considered as the dependent variable. To predict ROP, four machine learning algorithms were developed: Huber regression, Elastic Net, Ensemble, and Decision Tree. The models were then trained and tested on the collected data and evaluated using statistical metrics like Mean Absolute Percentage Error (MAPE), Coefficient of Determination (R^2), Root Mean Square Error (RMSE) and Explained Variance Score (EVS). The final model was the one with the best predictive accuracy and input-output consistency. The unique characteristics of this study are the simultaneous consideration of four important parameters and more importantly the inclusion of electrical conductivity of the drilling fluid as a new input variable for ROP modeling. Electrical conductivity of drilling fluid is a new perspective to analyze the interactions between rock properties, drilling conditions and bit performance.

2. Materials and Methods

This study was conducted in two main phases. In the first phase, field and laboratory investigations were carried out to establish the statistical population. In the second phase, supervised machine learning models were employed to identify the most accurate model for predicting the rock penetration rate. A flowchart illustrating the stages of this study is presented in Figure 1.

2.1. Site investigations and laboratory tests

The validity of the rock mechanics test results hinges on the selection of rock samples, since the accuracy of the rock mechanics tests is directly related to the quality of the samples and their representativeness relative to the intact rock mass. In this study, field studies were performed in quarries and stone-processing plants after preliminary studies on rock drilling and cooling-lubrication fluids. In this stage, seven building stone quarries were visited and seven rock blocks were finally selected as the research specimens which included three hard samples (granite) and four soft samples (marble and travertine). The collected samples were carried to the rock mechanics laboratory, cut to a standard size and prepared for physical, mechanical and drilling performance tests. Figure 2 illustrates the geographical locations of the quarries from which the rock samples were obtained.

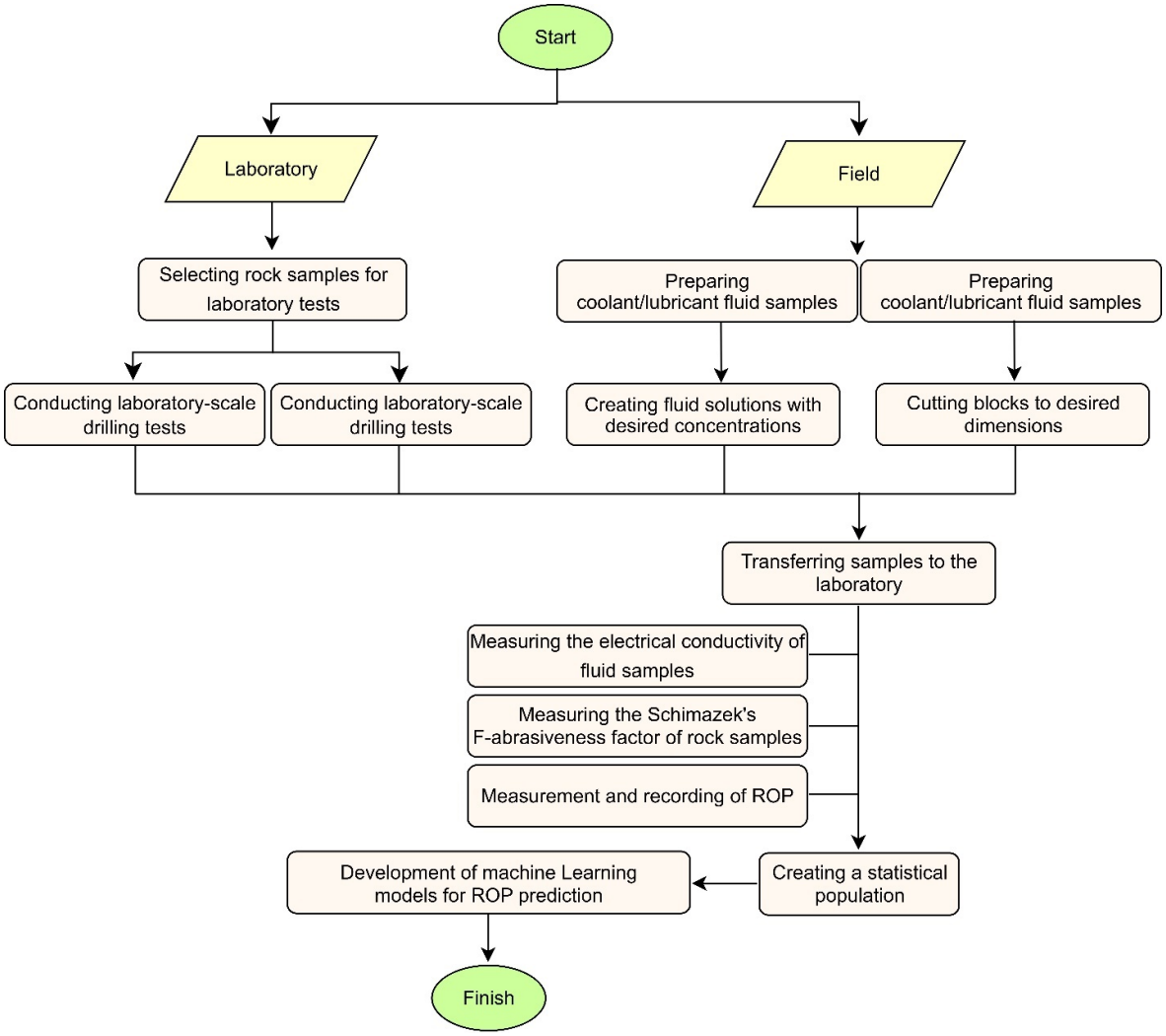

Figure 1. Flowchart of the research process

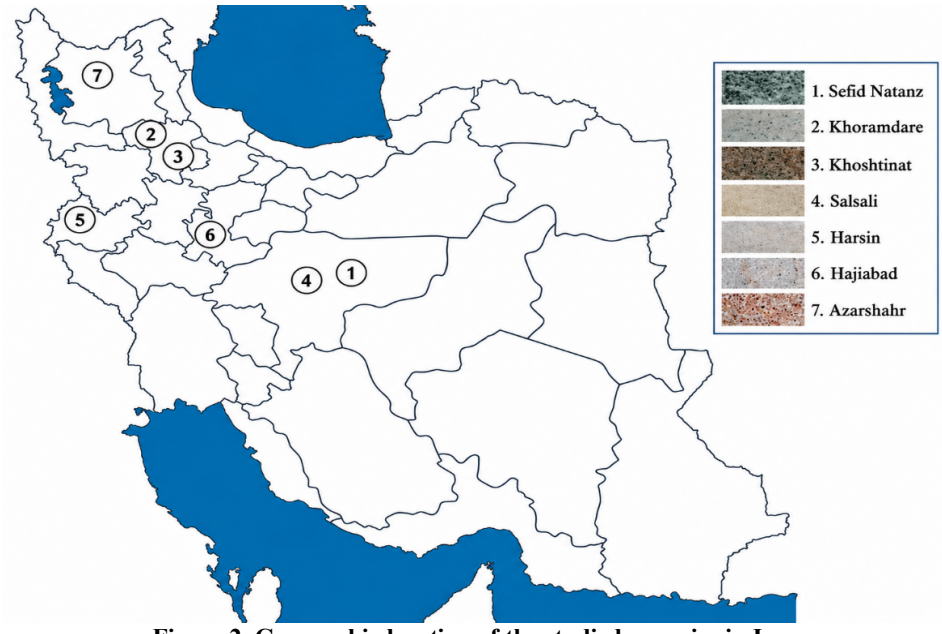

Figure 2. Geographic location of the studied quarries in Iran

In this study, one of the most important indirect indicators for assessing the abrasive characteristics of rocks-Schimazek's F-abrasiveness Index (SF_a)-was utilized. Rock abrasivity is a key parameter in the optimal design and cost estimation of drilling and cutting processes, as increased abrasion not only raises operational expenses but also reduces the life time of cutting tools and drill bits. Among various direct and indirect assessment methods, the SF_a index has been widely applied in geotechnical and mining projects due to its computational simplicity and strong correlation with the mechanical properties of rocks [60].

This index was first introduced by Schimazek and Knatz in 1970 and has since been recognized in numerous studies as a reliable criterion for estimating the abrasive potential of rocks. The general form of the equation for calculating this index is presented in Equation (1) [61].

$$SF_a = \frac{EQC \times G_s \times BTS}{100} \quad (1)$$

Where are:

SF_a – Schimazek's F-abrasiveness factor (*N/mm*),

EQC – the equivalent quartz content percentage,

G_s – the average grain size (*mm*),

BTS – the indirect Brazilian tensile strength (*MPa*).

The Schimazek's F-abrasiveness plays a key role in optimizing drilling operations. This index enables the prediction of both tool wear rate and ROP, providing a valuable basis for selecting appropriate drilling tools and determining optimal operational conditions. Its primary importance lies in its contribution to reducing operational costs and enhancing the overall efficiency and productivity of the drilling process [62, 63].

To assess the abrasive characteristics, two perpendicular thin sections were prepared from each rock sample and examined under a polarizing microscope. The petrographic analysis enabled precise identification of the rock's texture and mineral composition, from which the EQC and G_s were determined. Subsequently, to measure the BTS, a series of tests was conducted on core samples with a diameter of 54 *mm* and a height-to-diameter ratio of 0.5. Sample preparation followed ASTM D3967 standards, involving precise cutting and machining to achieve uniform, standard dimensions. This method is widely used for determining rock tensile strength due to its simplicity, reliability, and ease of sample preparation [64]. Finally, the data obtained from the petrographic analysis (EQC and Gs) and the BTS were used as primary inputs in Equation (1) to calculate the SF_a for each sample. The final results of these calculations are presented in Table 1, showing the Schimazek abrasiveness index values for the studied rock samples.

Table 1. Calculated Schimazek's F-abrasiveness index (SF_a) values for the studied rock samples.

Sample no.	Commercial name	Rock type	BTS (*MPa*)	EQC (%)	G_S (*mm*)	Sf_a (*N/mm*)
1	Sefid Natanz	Granite	11.28	55	13.54	2.16
2	Khoramdare	Granite	9.6	41	11.2	3.1
3	Khoshtinat	Granite	8.3	32.2	10.26	3.9
4	Salsali	Marble	6.3	3.2	0.105	0.52
5	Harsin	Marble	6.8	3.6	0.136	0.55
6	Hajiabad	Travertine	5.6	2.6	0.125	0.85
7	Azarshahr	Travertine	3.4	2.8	0.122	1.01

The appropriate selection of drilling fluids plays a fundamental role in enhancing drilling efficiency and increasing the reliability of results, as the type and physical properties of the fluid including viscosity, concentration, and thermal conductivity can directly affect tool performance and the overall quality of drilling operations. In this study, the fluids were selected based on preliminary investigations and consultations with drilling experts to cover a wide range of physical behaviors. These fluids primarily possessed cooling and lubricating properties, which are critical for reducing wear and improving drilling efficiency.

Considering the importance of electrical conductivity as a key indicator in evaluating drilling fluid performance, five cooling-lubricating fluids at specified concentrations, along with a base fluid (water), were selected and transferred to the laboratory for further testing. At this stage, the electrical conductivity of each fluid was measured precisely, and the resulting data were recorded. Table 2 presents the results of the laboratory

measurements for the electrical conductivity of the fluids under study.

The obtained data indicate that variations in the type and concentration of additives lead to significant changes in the electrical conductivity of the fluids, which can, in turn, influence heat transfer, lubrication, and ultimately the abrasive behavior of the tools. Therefore, the correct selection of drilling fluid and the control of properties such as EC play a crucial role in optimizing the drilling process and providing a solid foundation for reducing operational costs and predicting the ROP in the subsequent stages of this study.

Table 2. Electrical conductivity of the drilling fluids used in this study

Sample no.	Commercial name	Concentration of the additive in the water	EC (*μs/cm*)
1	Water	-	386
2	Soap water	1/60	147.3
3	Soap water	1/120	102.5
4	Syncool	1/100	166.4
5	Syncool	1/120	152.8
6	Boron nitride powder	5/20	238.9

Note: All concentrations are in liters of additive solved in the specified liters of water

2.1.1. Rock Drill Machine Construction

The penetration drilling rate was evaluated using a designed and constructed laboratory-scale drilling apparatus to obtain consistent and reliable data. Heterogeneity caused by differences in equipment, operation conditions and geological characteristics often affects the field data, which increases uncertainty and reduces the reliability of results. This apparatus is designed to control the laboratory conditions accurately, allows the reproducible data generation, reduces the cost, improves the safety and increases the measurement accuracy significantly.

It consists of two main parts: the drilling section and the processing block. The drill section consists of a rotary motor, a weighted loading system, a sample-holding platform, stabilizing clamps and a flushing system. The loading is affected by weights, which are attached by means of a chain and a gear mechanism, while the samples are placed on a circular platform having a vertical mobility and are fixed by clamps so as to prevent any displacement during the testing. The flushing system consists of a fluid reservoir, pump and adjustable nozzle. This allows for accurate control of flow rate and precise delivery of coolant and lubricant to the drill bit.

The processing unit consists of an electrical control panel equipped with a central processor and a touchscreen display, responsible for controlling operational parameters (such as rotation speed and flushing system) and recording real-time data on ROP and fluid flow.

A schematic overview of the apparatus is presented in Figure 3. This laboratory-scale drilling system provides an appropriate platform for investigating the effects of various parameters on the drilling process, generating precise and reliable data for developing ROP predictive models and optimizing drilling operations.

In this study, rock samples were prepared in standard dimensions of 10×10×15, 10×10×20, and 10×10×30 cm^3 and mounted on the circular platform of the drilling apparatus. To stabilize their position and prevent any displacement during the operation, the samples were firmly secured using specialized clamps. The drilling process was performed using blade-type drill bits equipped with tungsten carbide tips and diamond-tipped edges. These drill bits are specifically designed for hard rocks such as granite and have a diameter of 10 *mm*. To eliminate interaction effects between drilled holes, the spacing between them was set at twice the drill diameter (approximately 30 *mm*), and the holes were aligned along the central axis of each sample to ensure uniform conditions and comparability of results. Figure 4 presents an overall view of the drilling apparatus and the rock samples used in the rock mechanics laboratory.

To determine the optimal conditions, a series of preliminary trial-and-error experiments was conducted on samples with varying hardness. Ultimately, rotational speeds of 1190, 1057, and 933 *rpm* combined with axial loads of 1330, 1140, 930, and 755 *N* were selected for hard rocks, while speeds of 933, 845, 720, and 610 *rpm* with loads of 1140, 930, 755, and 570 *N* were chosen for soft rocks.

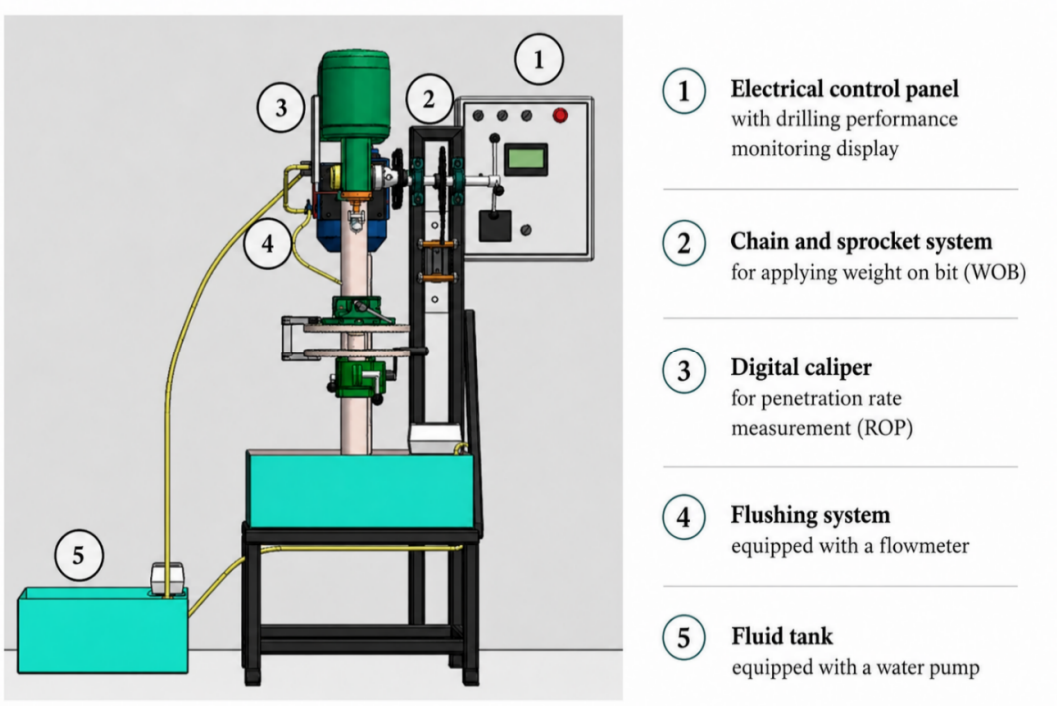

Figure 3. General view of the laboratory drilling apparatus designed in this study



Figure 4. General view of the drilling apparatus and rock samples

To evaluate the effect of cooling and lubricating fluids, experiments were first performed on medium and soft hardness samples, comparing the performance of water with five different fluids. The results indicated that fluids 3 and 4 showed the greatest performance improvement compared to water; therefore, these two fluids were selected for further tests on hard rock samples.

In total, 492 experiments were designed and executed. The main goal of this dataset was to investigate the relationships between rock mechanical parameters, fluid physical properties,

and operational parameters of the drilling apparatus for accurate prediction of ROP. The findings provided a valuable basis for analyzing drilling behavior across seven rock samples (including both hard and soft types) under conditions using three fluids for hard rocks (water plus two cooling/lubricating fluids) and six fluids for soft rocks (water plus five cooling/lubricating fluids).

This diversity in rock types and fluid selection enabled coverage of a wide range of drilling conditions and behaviors, creating a robust framework for analyzing the interactive effects of rock mechanical properties, fluid characteristics, and operational conditions on drilling performance.

2.2. Prediction methods

The selection of the four supervised learning algorithms, namely Huber Regression, Elastic Net, Ensemble (AdaBoost), and Decision Tree, was guided by three strategic considerations. First, fundamentally distinct modeling paradigms were intentionally represented, including robust regression (Huber), regularized linear methods (Elastic Net), boosting-based ensemble learning (AdaBoost), and tree-based nonparametric regression (Decision Tree). Such diversity enabled a comparative assessment across different modeling approaches. Second, given the relatively modest dataset size of 492 samples, preference was given to models with lower complexity, as these approaches are generally less susceptible to overfitting when training data are limited. Third, the selected algorithms provide a reasonable balance between predictive capability and interpretability, which facilitates the analysis of the relationships between the input variables and the rate of penetration.

It is acknowledged that advanced machine-learning methods such as Random Forest, XGBoost, LightGBM, Support Vector Regression, and Artificial Neural Networks have demonstrated strong predictive performance in rate-of-penetration (ROP) modeling [43, 56] .However, the primary objective of the present study was to compare representative algorithms from different modeling categories and to identify an appropriate predictive framework for the investigated dataset rather than to provide an exhaustive comparison of all available machine-learning techniques. Previous studies have also shown that tree-based and ensemble-learning approaches can achieve competitive performance in ROP prediction problems, particularly when dataset size is limited and the relationships between variables exhibit nonlinear behavior [24] .Therefore, the inclusion of additional algorithms may provide further insights into model performance and is recommended for future studies.

2.2.1. Huber regression method

The Huber regression method is considered a powerful technique for handling outlier data. By assigning lower weights to outliers, this method reduces their influence on the predictive model. In fact, Huber regression is classified as a robust regression technique. While conventional regression methods generally perform well under their underlying assumptions, they can yield unreliable results when these assumptions are violated. In contrast, Huber regression can provide reliable outcomes even when the standard assumptions are not satisfied.

The foundation of the Huber method is based on the Huber loss function, which combines the advantages of both the mean squared error and the mean absolute error. Specifically, the Huber function uses the squared error when the difference between the predicted and actual values is small, and the absolute error when this difference is large. A parameter, δ, is defined for the Huber function, which determines the threshold: for residuals smaller than δ, the squared error is applied, while for residuals larger than δ, the absolute error is used. The Huber loss function can be expressed mathematically as Equation (2) [65].

$$Huber = \begin{cases} \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}(y_i-\hat{y}_i)^2 & |y_i-\hat{y}_i| \le \delta \\ \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}\delta\left(|y_i-\hat{y}_i|-\frac{1}{2}\delta\right) & |y_i-\hat{y}_i| > \delta \end{cases} \quad (2)$$

Whera are:

y_i – the observed value of sample i,

$\hat{y}_i$ – the predicted value of sample i,

$y_i - \hat{y}_i$ – the residual of sample i,

n – the number of observations,

δ – the threshold parameter controlling transition between quadratic and linear regions.

2.2.2. Elastic network method

The elastic net regression was introduced by Zou and Hastie in 2005 [66]. This method is considered a hybrid approach for predictive modeling. By simultaneously leveraging the regularization functions of both Lasso and Ridge regression techniques, the elastic net demonstrates high efficiency in developing predictive models.

When dealing with high-dimensional data, phenomena such as overfitting and multicollinearity among independent variables may arise during the training and testing processes. The elastic net technique effectively addresses these issues through appropriate selection of the regularization function. The algorithm enhances the overall modeling process by adding penalty terms to the ordinary least squares objective function.

Specifically, this method incorporates two compensatory terms, α (alpha) and λ (lambda), which correspond to the L1 (Lasso) and L2 (Ridge) regularization functions, respectively, and serve as the hyperparameters of the elastic net model. The elastic net model can be expressed mathematically as shown in Equation (3).

$$y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n + e \quad (3)$$

Where are:

y – the dependent variable,

b_0 – the intercept,

b_1 to b_n –the regression coefficients,

x_1 to x_n – the independent variable,

e – the error term.

In the elastic net model, the objective is to minimize the loss function presented in Equation (4).

$$\text{RSS} + \lambda \times [(1-\alpha) \times \|\beta\|_2 + \alpha \times \|\beta\|_1] \quad (4)$$

Where are:

RSS – the residual sum of squares,

λ – the regularization parameter,

β – the vector of coefficients,

α – the parameter that determines the balance between the L1 and L2,

$\|\beta\|_1$ – the L1 norm of β,

$\|\beta\|_2$ – the L2 norm of β.

Due to its hybrid nature, the elastic network method exhibits high robustness against factors that may degrade the performance of predictive models. It effectively accounts for differences in data scale, showing minimal sensitivity to such variations. When faced with independent variables exhibiting high collinearity, the method provides reliable results. Overall, compared to other linear predictive models, the elastic network method demonstrates superior performance.

2.2.3. Ensemble method

Ensemble models aim to improve predictive performance by combining multiple individual models. The models used within an ensemble framework are referred to as base models. By leveraging base models, ensemble techniques address issues such as high variance in input data, low accuracy of the final model, the presence of noise, and the overreliance of the model on specific input features. By integrating the outputs of base models, ensemble methods can both control and manage the overall error of the final model and achieve high reliability during model generalization. To attain acceptable error levels and high generalization capability, various techniques-collectively known as meta-algorithms-are employed. The four major techniques commonly used in ensemble methods are bagging, boosting, blending, and stacking. When applying ensemble algorithms for predictive modeling, combining the outputs of multiple base models requires specific aggregation methods. The three primary aggregation approaches are max voting, averaging, and weighted averaging. A schematic overview of the ensemble technique is presented in Figure 5 [67].

The ensemble model implemented in this study utilizes the AdaBoost.R2 algorithm (Adaptive Boosting for Regression), which differs from the AdaBoost.M1 variant originally designed for classification. The base learner architecture consists of decision tree regressors with a maximum depth of 3 (i.e., simple trees with a maximum of 8 leaf nodes). This depth was intentionally selected to ensure the creation of 'weak learners' that perform only marginally better than random chance, with an expected R^2 in the range of 0.15-0.25. The sequential boosting strategy is formulated as follows. In the first iteration (t = 1), uniform weights are assigned to all training samples according to Equation (5).

$$w_i^1 = \frac{1}{n} \quad (5)$$

Subsequently, a decision tree regressor is fitted to the weighted training data, and the weighted absolute error (ϵ_t) is computed. The contribution of each weak learner to the final ensemble prediction is then determined through the learner weigh (α_t), which is calculated using Equation (6).

$$\alpha_t = \frac{1 - \varepsilon_t}{\varepsilon_t} \quad (6)$$

which determines the contribution of the t-th weak learner to the final ensemble prediction. The sample weights for iteration $(t+1)$ are then updated according to Equation (7).

$$w_i^{(t+1)} = w_i^{(t)} \times exp\left(\alpha_t \times error_{i,t}\right) \quad (7)$$

where $error_{i,t}$denotes the normalized absolute error for the sample i at iteration t. This weighting scheme ensures that subsequent weak learners focus more heavily on samples with larger prediction errors. The final prediction is computed as the weighted median of the weak learners' outputs; compared to a weighted average, this approach provides greater robustness against outliers.

The hyperparameters of the ensemble were optimized using a Grid Search approach as follows: the number of estimators was set to 50 (representing the maximum number of boosting stages, with an early stopping mechanism implemented if the test error failed to decrease for 10 consecutive iterations); the learning rate (η) was set to 1.0 (the shrinkage factor applied to each α_t; a value of $\eta = 1.0$ implies no shrinkage, which was deemed appropriate given the shallow depth of 3 for the base learners); and the loss function was specified as 'linear'-the error metric used for updating sample weights-where 'linear' corresponds to the absolute error, distinguishing it from 'square' or 'exponential' loss functions.

Consequently, the final ensemble model is obtained by combining the outputs of all weak learners through the weighted median formulation given in Equation (8).

$$F(x) = median_{t=1}^{50}[\alpha_t \times f_t(x)] \quad (8)$$

where $f_t(x)$denotes the prediction of the t-th weak decision tree.

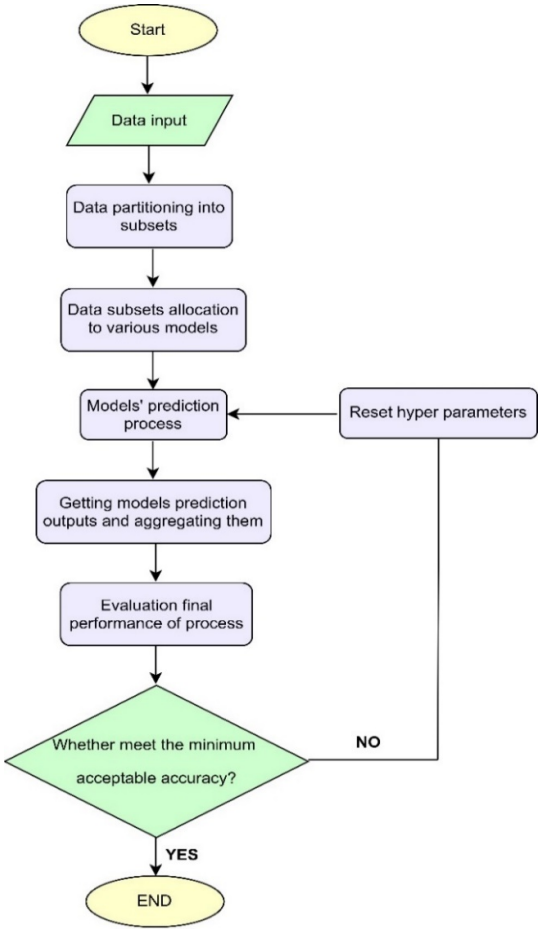

Figure 5. General schematic of the ensemble technique

2.2.4. Decision tree method

Supervised regression decision trees are built based on a tree-like structure. This technique partitions the dataset into smaller subsets. The final outcome is a hierarchical tree consisting of decision nodes and leaf nodes. Each decision node contains two or more branches, each representing a specific feature of the data. Leaf nodes indicate the decision based on numerical targets. The topmost decision node is called the root node.

The primary algorithm for constructing a decision tree is ID3, introduced by J.R. Quinlan. ID3 uses a greedy top-down search strategy, exploring the space of possible branches. In regression tasks, decision trees replace variance measures with standard deviation to guide tree construction.

Decision trees are among the most widely used pattern recognition techniques in data mining. Figure 6 illustrates a schematic representation of the key concepts and overall structure of the decision tree method [68].

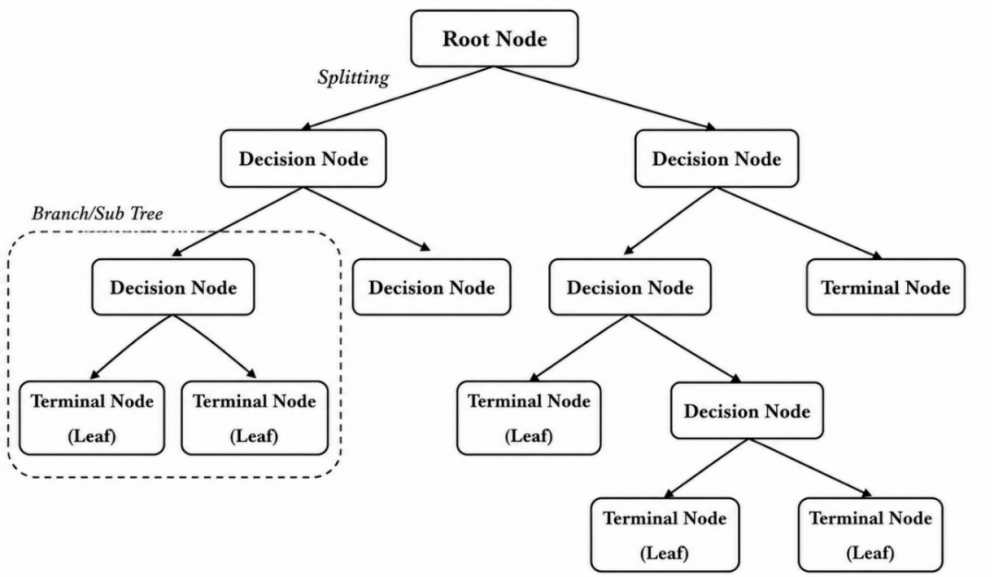

Figure 6. General schematic of the key concepts and overall structure involved in the decision tree method

2.3. Methods for evaluating the performance quality of models

Prior to model development, the complete dataset comprising 492 laboratory measurements was randomly partitioned using stratified random sampling to preserve proportional rock-type representation across all subsets. The training set consisted of 70% of the data (n = 344 observations) and was used for model calibration and hyperparameter optimization. The testing set comprised the remaining 30% (n = 148 observations) and was reserved exclusively for final model evaluation. The 70:30 partitioning strategy was selected to provide sufficient training data for effective pattern identification while maintaining a robust testing set for unbiased performance assessment. Reproducibility was ensured through the use of a fixed random seed (42). For model validation, k-fold cross-validation (k = 5) was implemented during hyperparameter tuning for the Huber, Elastic Net, and Ensemble models to reduce the risk of overfitting and improve generalization capability. Given the inherent susceptibility of Decision Tree models to overfitting, additional cross-validated pruning was performed through cost-complexity parameter tuning using 5-fold cross-validation.

2.3.2. Mean absolute percent error

One of the statistical parameters used to evaluate model accuracy, which is highly regarded in statistics, is the Mean Absolute Percentage Error (MAPE). This metric is based on a loss function framework. By providing an objective interpretation of the prediction errors, it serves as a powerful tool for comparing the performance of different models. Considering the contributing factors in its calculation and formulation, MAPE not only assesses the accuracy and reliability of a model but can also provide insights into its stability. The error is obtained by calculating the absolute difference between the predicted values and the measured (or actual) values. Equation (9) presents the formula for calculating the mean absolute percentage error.

$$M = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - P_t}{A_t}\right| \quad (9)$$

Where are:

M – mean absolute percentage error

n – number of times the summation iteration happens

A_t – actual value

P_t – predicted value

2.3.2. Determination coefficient

This evaluation metric essentially represents the portion of the variance in the dependent variable that can be explained or predicted by the independent variables. The coefficient of determination (R^2) provides more objective and comprehensive information compared to other evaluation metrics. This is because R^2 has a fixed range between 0 and 1, whereas other metrics, such as Mean Absolute Error (MAE) or Mean Relative Absolute Error (MRAE), can fluctuate over an arbitrary range. Moreover, R^2 demonstrates greater

stability when confronted with unsuitable or noisy data compared to other evaluators. Fundamentally, the coefficient of determination quantifies the quality of fit of the proposed model to the actual observed data. Equation (10) can be used to calculate R^2.

$$R^2 = 1 - \frac{RSS}{TSS} \quad (10)$$

Where are:

R^2= Coefficient of determination

RSS = Sum of squares of residuals

TSS = Total sum of squares

2.3.3. Root mean square error

The Root Mean Square Error (RMSE) is one of the key statistical parameters used to evaluate the performance quality of predictive models. In general, smaller RMSE values-approaching zero-indicate higher model accuracy and better predictive capability. However, due to the non-fixed scale and dependency of RMSE on the range of the target variable, it is not always suitable for direct comparison of model performance across different datasets. RMSE is calculated as the square root of the average of the squared differences between the predicted and actual values. The mathematical expression for RMSE is presented in Equation (11).

$$MSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(A_t - P_t)^2} \quad (11)$$

Where are:

MSE – mean squared error

n – number of data points

A_t – actual value

P_t – predicted value

2.3.4. Explained variance score

The Explained Variance Score (EVS) describes the degree of dispersion of the prediction errors generated during the modeling process. In other words, the explained variance score complements the unexplained variance, which corresponds to the residual errors. This metric reflects how well the predictive model captures the variability of the actual data by quantifying the spread of the prediction errors around the observed values. The explained variance score can be calculated using Equation (12).

$$\text{EVC} = 1 - \frac{var(A_t - P_t)}{var(A_t)} \quad (12)$$

Where are:

Var – variance

A_t – actual data

P_t – predicted data

2.3.5. Target value

Since multiple evaluation metrics were employed to more accurately assess the performance quality of the models, it is not possible to reliably identify the best-performing model solely by comparing the individual values of these metrics. To address this issue, a composite function, referred to as the objective function, was developed by integrating the evaluation parameters. The objective function is formulated such that lower output values correspond to better model performance. It should be noted that the arrangement and weighting of the evaluation parameters in constructing the objective function are not unique-alternative combinations can also be considered. Equation (13) presents the formula used to calculate the objective function based on the selected evaluation metrics.

$$Target\,func. = MAPE + RMSE + |1 - R^2| + |1 - EVS| \quad (13)$$

The objective function was formulated to integrate multiple complementary performance metrics into a single scalar value, thereby enabling unambiguous ranking of the developed models. The four selected evaluation metrics, namely MAPE, RMSE, R^2, and EVS, were chosen because distinct aspects of predictive performance are captured by each measure. The MAPE metric is used to quantify the relative magnitude of prediction errors independently of variable scale, whereas RMSE imposes a quadratic penalty on large prediction errors and is therefore particularly sensitive to performance degradation caused by outliers. The coefficient of determination (R^2) measures the proportion of variance explained by the model, while EVS evaluates the extent to which the variability of the target variable is reproduced by the predictive model. The objective function

presented in Equation (13), expressed as Target Function = MAPE + RMSE + $|1 - R^2| + |1 - EVS|$, was established according to three guiding principles. First, all components were transformed such that lower values consistently indicate superior predictive performance. Because MAPE and RMSE inherently satisfy this condition, no transformation was required for these metrics. In contrast, R^2 and EVS, which typically range between 0 and 1 with higher values representing better performance, were converted to a minimization framework through the transformations $|1 - R^2|$ and $|1 - EVS|$. Second, weighting coefficients were not assigned to the individual components, reflecting the assumption that each metric contributes equally to the overall assessment of model performance in the absence of a theoretically or empirically justified basis for differential weighting. Third, normalization was not applied because the selected metrics were considered sufficiently compatible for aggregation: MAPE is dimensionless, RMSE is expressed in the same units as the target variable (mm/s), and the transformed R^2 and EVS terms are dimensionless quantities. Consequently, the summation of these four components yields an objective-function value that can be interpreted as a composite performance-error index, with lower values indicating superior overall predictive capability.

2.4. Statistical Characterization of the Dataset

The dataset (n = 492) was examined with respect to its distributional characteristics, potential outliers, class balance, and variability among rock types. For the target variable (ROP), a distribution characterized by mild positive skewness (skewness = 0.67) and moderate kurtosis (2.9) was observed. Normality was evaluated using the Shapiro-Wilk test, for which a test statistic of W = 0.94 and a p-value of 0.072 were obtained. Because the null hypothesis of normality was not rejected at $\alpha = 0.05$, the application of parametric statistical methods for between-group comparisons was considered appropriate.

Potential outliers were identified using the interquartile range (IQR) method with a threshold of 1.5 × IQR. A total of 14 observations (2.8% of the dataset) were flagged as potential outliers. Subsequent examination revealed that these observations corresponded to valid experimental conditions, including elevated ROP values associated with travertine specimens under lubricated drilling conditions, rather than measurement or recording errors. Consequently, all observations were retained in the dataset. The influence of these observations on model performance was further evaluated, and it was found that their exclusion resulted in negligible changes in predictive accuracy ($\Delta R^2 < 0.01$ for all models), indicating that the reported results were not driven by outlier effects.

With respect to rock-type distribution, the dataset consisted of 204 granite observations (41.5%), 162 marble observations (32.9%), and 126 travertine observations (25.6%). This degree of imbalance was considered modest, as the ratio between the largest and smallest groups was 1.62:1, which falls within the range generally regarded as acceptable for regression-based modeling applications.

Variability in ROP among the rock types was assessed using one-way analysis of variance (ANOVA). Significant differences were detected among the groups ($F(2,489) = 28.4$, $p < 0.001$). Subsequent Tukey HSD post hoc analysis demonstrated that all pairwise comparisons were statistically significant at $\alpha = 0.05$. Specifically, the mean ROP observed for granite (1.42 mm/s, SD = 0.47) was significantly lower than the corresponding values for marble (2.31 mm/s, SD = 0.52) and travertine (2.85 mm/s, SD = 0.61). These results provide statistical evidence of substantial between-rock-type variability and support the inclusion of rock-related properties, represented through Schimazek's abrasivity index (SF_a), as predictive inputs within the modeling framework.

Finally, the coverage of the input space was evaluated through convex-hull volume analysis. The experimental observations were found to occupy approximately 72% of the theoretical input space defined by the ranges of the predictor variables, namely SF_a (0.52–3.9 N/mm), EC (102.5–386 µs/cm), WOB (570–1,330 N), and BRS (610–1,190 rpm). This result indicates that adequate representation of the input domain was achieved and that only limited extrapolation would be required when predictions are generated within the investigated parameter ranges.

3. Results and discussion

3.1. Preprocessing

An essential prerequisite for discovering the underlying patterns in a dataset is to examine the data before initiating the overall modeling process. One of the key steps in identifying the governing model is to investigate the behavioral relationships among the input variables. A common negative phenomenon observed among input variables is

"multicollinearity". Multicollinearity occurs when the directions of variation of two or more variables are proportional and aligned-that is, when the variables are linearly dependent and can be expressed as a nearly perfect linear function of one another. In statistical literature, this condition is referred to as "high correlation among variables".

If model development is conducted on data affected by multicollinearity, two major issues may arise. First, the regression coefficients associated with the collinear input variables can vary significantly with each iteration of the modeling process, leading to "model instability". This instability implies that the model lacks reliability and exhibits minimal certainty. Second, multicollinearity reduces the "generalizability" of the model, as the instability in coefficients prevents consistent performance on new datasets.

To prevent such issues, it is crucial to examine the degree of correlation among the input variables. A common and effective technique for this purpose is the "correlation matrix", which presents pairwise correlation coefficients between input variables. This matrix provides a clear and intuitive visualization of the relationships and interdependencies among variables, allowing for easy interpretation. Accordingly, to assess the input data in terms of potential multicollinearity, the "correlation coefficient matrix" of the input parameters is presented in Figure 7.

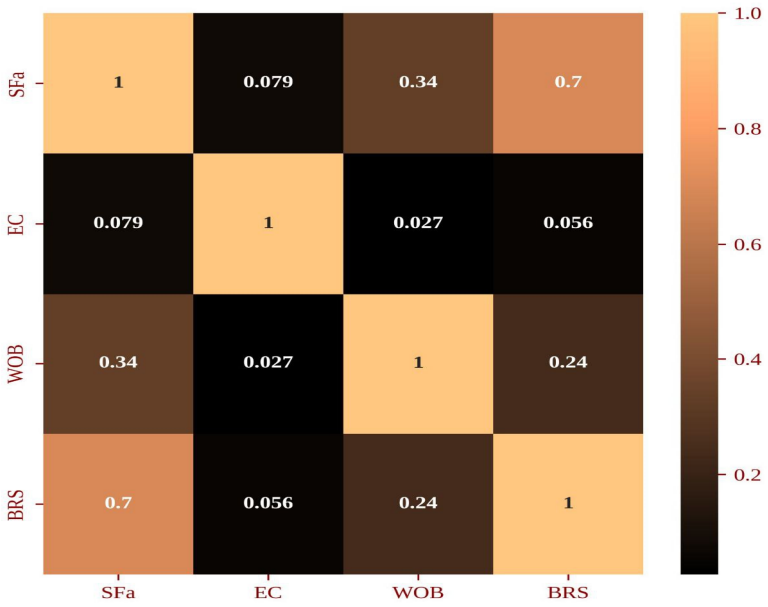

Figure 7. Correlation coefficient matrix of the input variables used in the models

In the correlation coefficient matrix of the input variables, the pairwise correlations among Schimazek's abrasivity index, bit load, EC of the drilling fluid, and bit rotation speed are presented. According to the data shown in Figure 7, the highest correlation coefficient (approximately 0.7) is observed between Schimazek's abrasivity index and the bit rotation speed. The remaining correlation coefficients fall within a much lower range, indicating weak inter-variable correlations.

A correlation value of 0.7 between two variables is generally considered acceptable. Although the degree of alignment between Schimazek's abrasivity index and the bit rotation speed is relatively higher compared to other variables, it remains within a range that is unlikely to cause significant multicollinearity or adversely affect the modeling results.

Prior to model development and hyperparameter optimization, all input variables, including Schimazek's abrasivity index (SF_a), electrical conductivity (EC), weight on bit (WOB), and bit rotation speed (BRS), as well as the target variable, rate of penetration (ROP), were standardized using the z-score transformation

Standardization was performed for three principal reasons. First, regularization terms employed by the Elastic Net and Huber Regression models, including L1 and L2 penalty functions, are sensitive to differences in variable scale. Without standardization, variables characterized by larger numerical magnitudes would exert a disproportionate influence on the regularization process, potentially resulting in biased coefficient estimates. Second, feature-scale homogeneity is generally assumed in distance-based calculations

incorporated within certain ensemble-learning algorithms, and standardization facilitates compliance with this assumption. Third, numerical stability during optimization is improved when variables are expressed on comparable scales, thereby enabling gradient-based optimization procedures to operate within more well-conditioned parameter spaces. The standardization parameters, namely the mean ((μ)) and standard deviation ((σ)), were calculated exclusively from the training dataset (n = 344) and were subsequently applied to transform both the training and testing datasets. This procedure was adopted to prevent information leakage from the testing dataset into the preprocessing stage and to ensure an unbiased evaluation of model performance. Although standardization is not required for Decision Tree models because their structure is invariant to monotonic transformations of the input variables, the same standardized dataset was provided to all models in order to maintain methodological consistency and ensure a fair comparison among the evaluated techniques.

Subsequently, each of the selected modeling techniques employed in this study was implemented to analyze and evaluate the modeling process and the predictive performance of the developed models. The first method examined is the Huber regression model, which includes four primary hyperparameters. Proper tuning of these hyperparameters within appropriate ranges can significantly influence the accuracy and robustness of the model's performance.

The Huber model hyperparameters and their optimized values are presented in Table 3. These optimal values were determined using a trial-and-error procedure, in which multiple ranges of hyperparameters were tested for each method, and the combinations yielding the best model performance were selected as optimal. Similarly, the optimized hyperparameter values for the Elastic Net, Ensemble, and Decision Tree techniques are reported in Tables 4, 5, and 6, respectively.

Table 3. Hyperparameters used in the Huber regression model

Parameters	Epsilon[1]	Maximum Iteration[2]	Alpha[3]	Tolerance[4]
Value	4	100	0.0001	0.00001

1. Epsilon: The parameter epsilon controls the number of samples that should be classified as outliers.
2. Maximum Iteration: Maximum number of iterations.
3. Alpha: Strength of the squared L2 regularization.
4. Tolerance: Iterations stop criteria.

Table 4. Hyperparameters used in the Elastic network model

Parameters	Alpha[1]	Maximum Iteration[2]	Selection[3]	Tolerance[4]	L1 Ratio
Value	0.1	1000	Cyclic	0.0001	0.1

1. Alpha: Constant that multiplies the penalty terms.
2. Maximum Iteration: Maximum number of iterations.
3. Selection: Random coefficients updating method.
4. Tolerance: Iterations stop criteria. 5. L1 Ratio: The Elastic Net mixing parameter.

Table 5. Hyperparameters used in the Ensemble model

Parameters	Ensemble Method[1]	Estimators Number[2]	Learning Rate[3]	Loss[4]
Value	Adaboost	50	1.0	Linear

1. Ensemble Method: Ensemble method name which is used.
2. Estimators Number: The maximum number of estimators at which boosting is terminated.
3. Learning rate: Weight applied to each regressor at each boosting iteration.
4. Loss: The loss function to use when updating the weights after each boosting iteration.

Table 6. Hyperparameters used in the Decision tree model

Parameters	Criterion[1]	Spliter[2]	Minimum Split [3]	Minimum Leaf[4]	Weight Fraction
Value	Squared Error	best	2	1	0.0

1. Criterion: The function to measure the quality of a split.
2. Splitter: The strategy used to choose the split at each node.
3. Minimum Split: The minimum number of samples required to split an internal node.
4. Minimum Leaf: The minimum number of samples required to be at a leaf node.
5. Weight Fraction: The minimum weighted fraction of the sum total of weights (of all the input samples) required to be at a leaf node.

As described in the methodology section, four statistical metrics including MAPE, R^2, RMSE, and EVS were employed to evaluate the performance quality of the predictive models developed in this study. At this stage of the research, these evaluation indices were calculated to assess the accuracy and robustness of each model. It should be noted that the overall modeling process involved dividing the dataset into training and testing subsets to ensure reliable evaluation. The calculated values of the quantitative performance indicators for each model are presented in Table 7.

Table 7. Calculated values of quantitative performance evaluation metrics for the models

Model	Mean Absolute Percentage Error	Determination Coefficient	Root Mean Squared Error	Explained Variance of Errors	Target Value
Huber	0.3075	0.5425	0.0483	0.5431	1.2702
Elastic Net	0.3349	0.4698	0.0520	0.4700	1.4471
Ensemble	0.2043	0.8278	0.0296	0.8406	0.5655
Decision Tree	0.1140	0.9170	0.0205	0.9193	0.2975

An examination of Table 7 provides valuable insights into the performance of the different predictive models. To facilitate comparison among the models, the performance results for each evaluation metric are illustrated in Figures 8 to 11. These figures allow for a clear visual interpretation of how each model performs relative to the others.

Figure 8 presents the results for the MAPE index. As shown, the performance of the four models demonstrates a clear distinction, with the Decision Tree model exhibiting the best performance by achieving the lowest mean absolute percentage error, indicating the highest prediction accuracy. In terms of MAPE, the Ensemble and Huber regression models occupy the next ranks, while the Elastic Net model, which yielded the highest MAPE value, demonstrates the weakest performance among the four.

Figure 9 illustrates the performance of the models based on the R^2. Given the significance of this metric, evaluating model performance through R^2 provides critical insights into how well each model explains the variance in the dependent variable. As observed in Figure 9, once again, the Decision Tree model demonstrates the best performance, achieving the highest R^2 value. In terms of explanatory power, the Ensemble model ranks second, while the Huber regression and Elastic Network models exhibit the lowest R^2 values, indicating comparatively weaker performance.

The next commonly used evaluation metric, RMSE, is presented in Figure 10. Once again, the Decision Tree model demonstrates the most accurate predictions, exhibiting outstanding performance in minimizing prediction errors. The pattern observed in the previous two metrics (MAPE and R^2) is repeated here: the Ensemble, Huber regression, and Elastic Network models follow in descending order of performance, with Elastic Network showing the least favorable results.

Finally, as shown in Figure 11, regarding the EVS, the Decision Tree model again provides the most accurate outputs. By achieving the highest EVS, it delivers the most reliable predictions. Similarly, for this metric, the Ensemble, Huber regression, and Elastic Net models rank subsequently in descending order of performance.

Overall, considering the performance of the four models across the previously discussed evaluation metrics, it is evident that the objective function, which integrates these metrics, should also reflect the superior performance of the decision tree method. It is important to note that the objective function is designed based on the arrangement of the aforementioned metrics, such that better model outputs correspond to lower objective function values. Accordingly, Figure 12 illustrates the objective function values for the models under study. The figure clearly shows that the decision tree-based model exhibits the best performance, followed by the ensemble, Huber, and elastic net models, respectively.

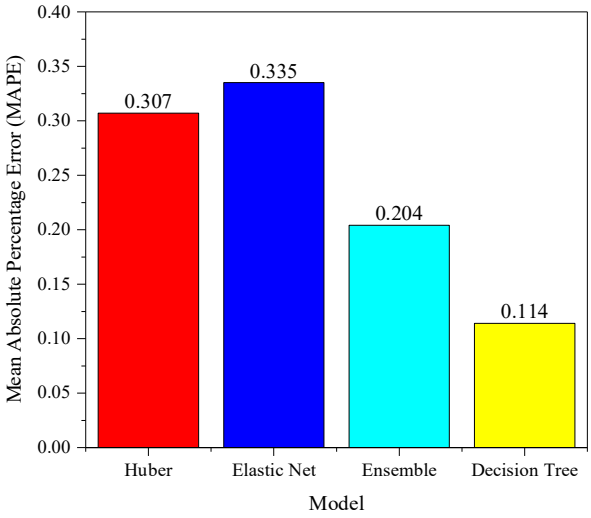

Figure 8. Performance of the models based on the MAPE

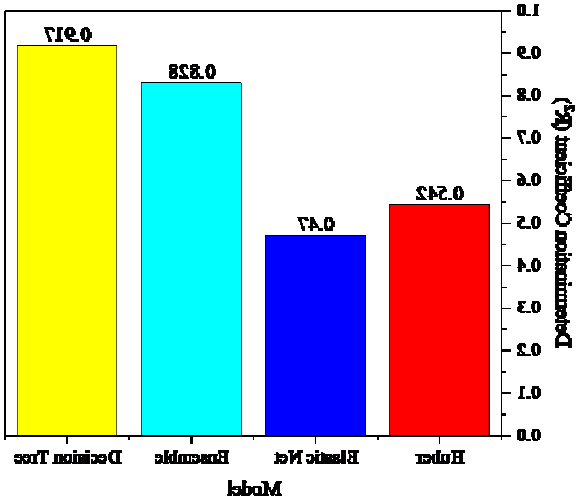

Figure 9. Performance of the models based on the R^2

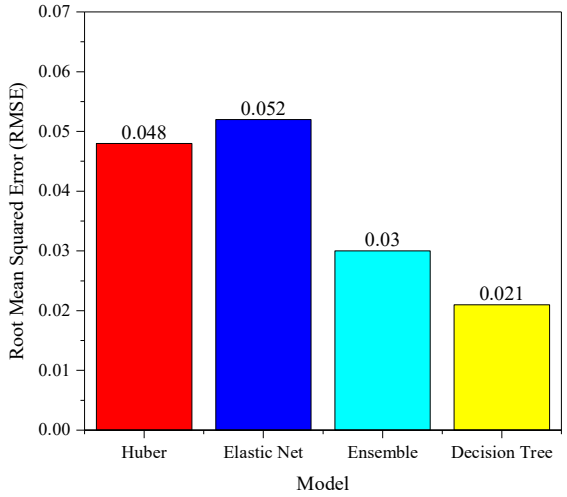

Figure 10. Performance of the models based on the RMSE

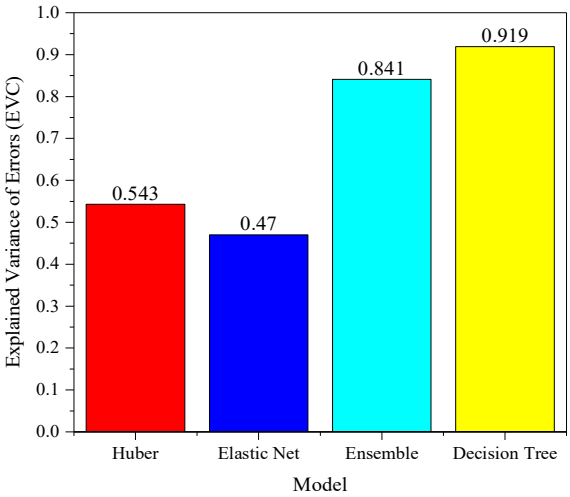

Figure 11. Performance of the models based on the EVC

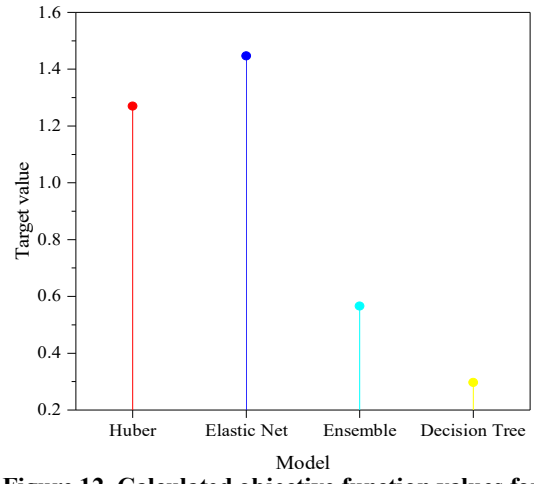

Figure 12. Calculated objective function values for the used models

In general, based on Figures 8 to 12, it can be concluded with confidence that among the models used to predict the drill bit penetration rate in rock, the decision tree-based model demonstrated the best performance. A graphical comparison of actual versus predicted data can serve as an additional indicator of the performance efficiency of machine learning models. The closer the predicted values are to the actual values, the higher the modeling accuracy. It is important to note that proximity is preferable to exact alignment; perfect alignment may indicate overfitting, which would reflect an unreliable predictive model.

Figures 13 to 16 illustrate this comparison for the Huber, Elastic Net, Ensemble, and Decision Tree models. In these plots, the horizontal axis represents the data indices, while the vertical axis

shows the penetration rate. Two types of lines are depicted: dashed lines indicate predicted values, and solid lines represent measured (actual) values. Predicted data points are marked with small squares, and actual data points with small circles.

The performance of each model in these plots can be evaluated based on the separation between the two lines: the greater the distance between the predicted and actual values, the poorer the model's performance. Conversely, the closer the two lines are, the more accurate and reliable the model is. According to this criterion, the decision tree model shows the closest alignment between predicted and actual values, further confirming its superior performance compared to the other models and reinforcing the high-quality predictive capability of the decision tree approach.

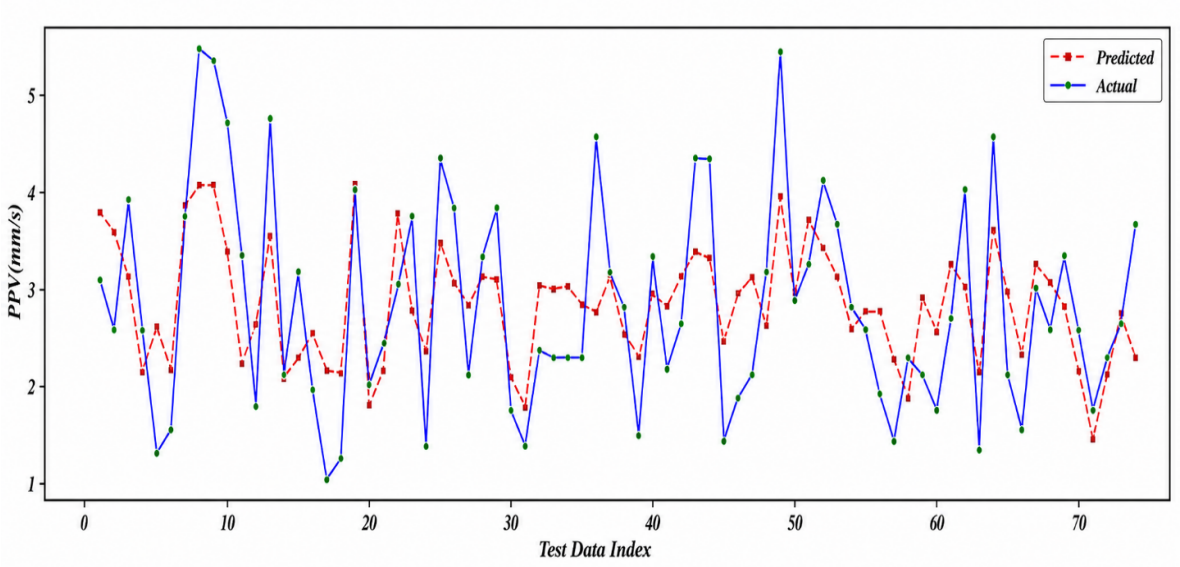

Figure 13. Time series comparison of the maximum predicted particle velocity versus the maximum actual particle velocity using the Elastic Net model

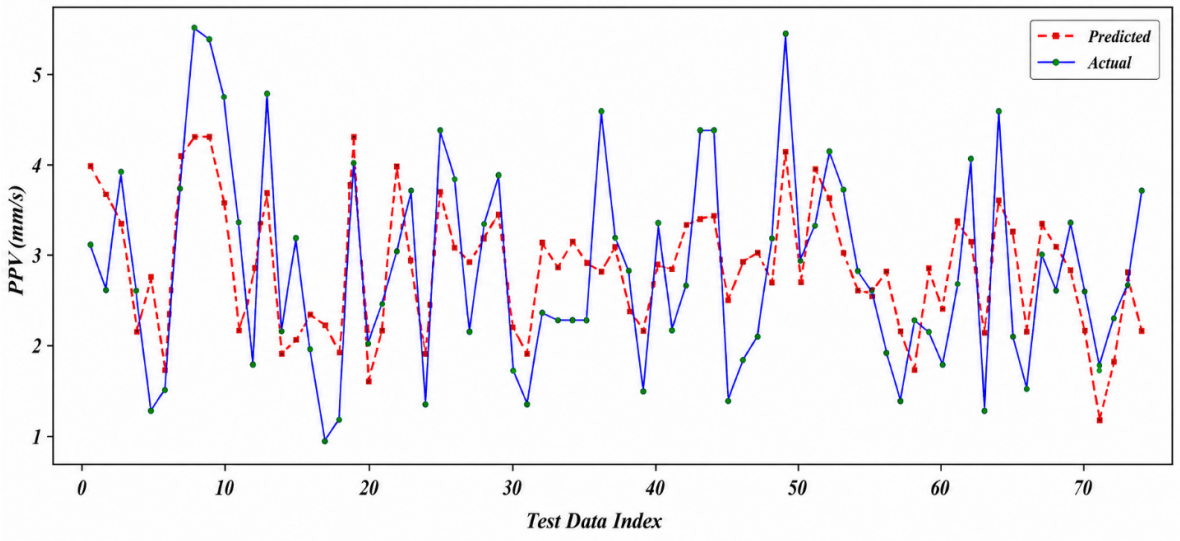

Figure 14. Time series comparison of the maximum predicted particle velocity versus the maximum actual particle velocity using the Huber model

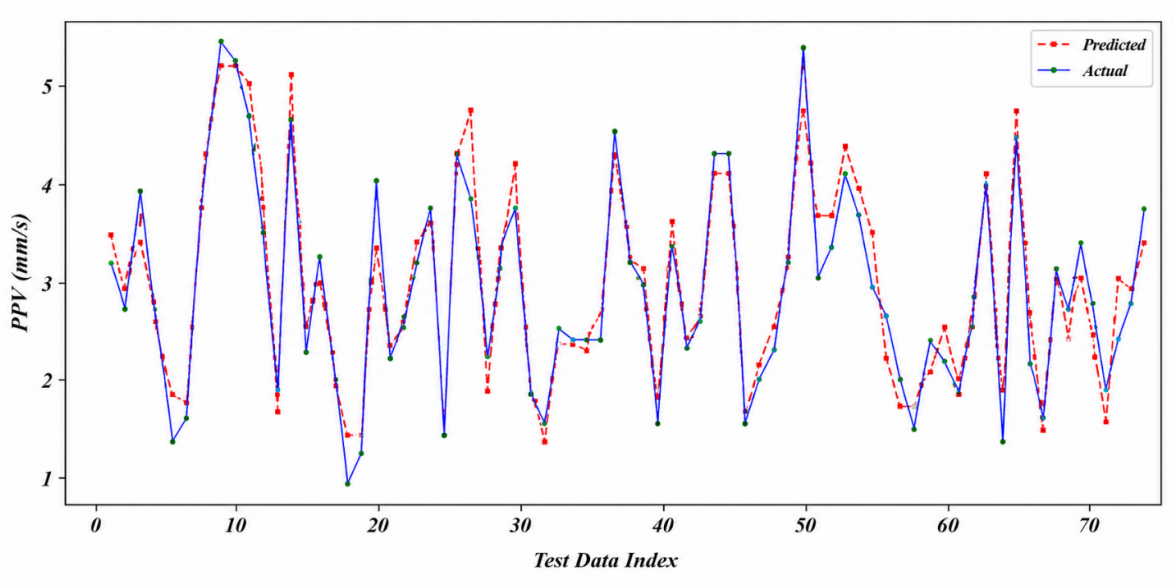

Figure 15. Time series comparison of the maximum predicted particle velocity versus the maximum actual particle velocity using the Ensemble model

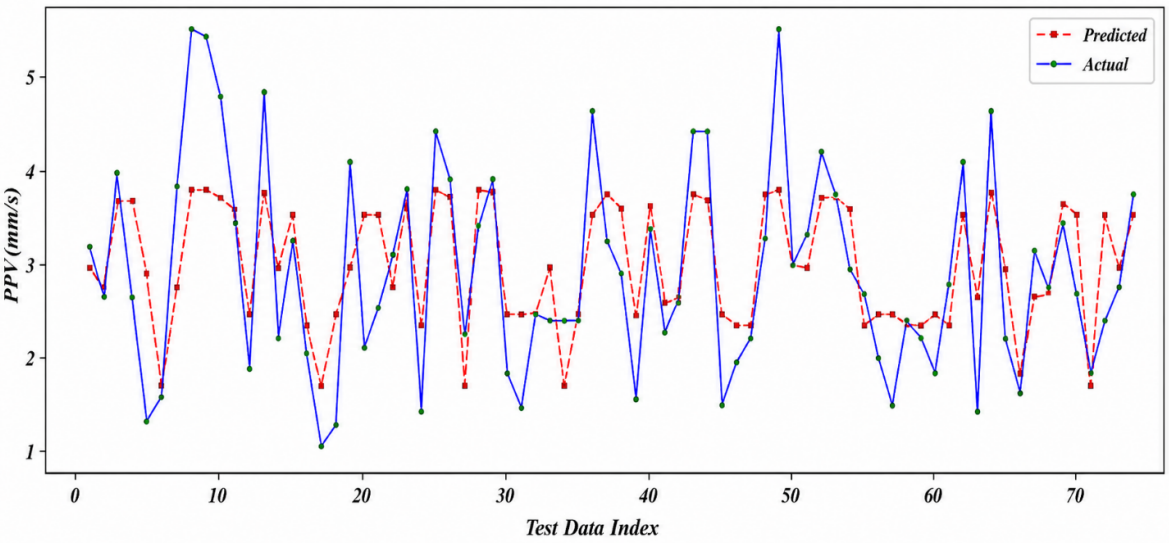

Figure 16. Time series comparison of the maximum predicted particle velocity versus the maximum actual particle velocity using the Decision tree model

To further evaluate the predictive performance of the models, residual error plots were generated in two different formats. In the first type of plot, the actual values (i.e., measured values) are represented on the vertical axis, while the predicted values are plotted on the horizontal axis. In this type of plot, as the predictive performance of the model improves, the actual and predicted values converge toward each other and cluster around the bisector of the first quadrant. In other words, a dense cluster of actual and predicted data points forms around the bisector. The greater the density and compactness of this cluster, the more reliable the model's performance is considered to be.

In the second type of plot, illustrated in Figures 17 and 18, the residual errors resulting from the model's predictions are plotted on the vertical axis, while the predicted values are shown on the horizontal axis. The structure of this plot is such that, for a poorly performing model, the points are widely scattered across the plot, indicating substantial deviation of the predicted values from the actual values. This dispersion reflects large residual errors and demonstrates low predictive accuracy. Conversely, for a high-performing and

reliable predictive model, the points cluster closely around the horizontal line passing through zero on the vertical axis, forming a compact longitudinal cluster. This pattern indicates low residual errors for each predicted value and suggests that the errors are close to zero. Figures 17 and 18 illustrate these residual and predicted-versus-actual value plots. Examination of these plots confirms that the decision tree-based model exhibits the best performance, as evidenced by the highest density of clusters in its respective plots.

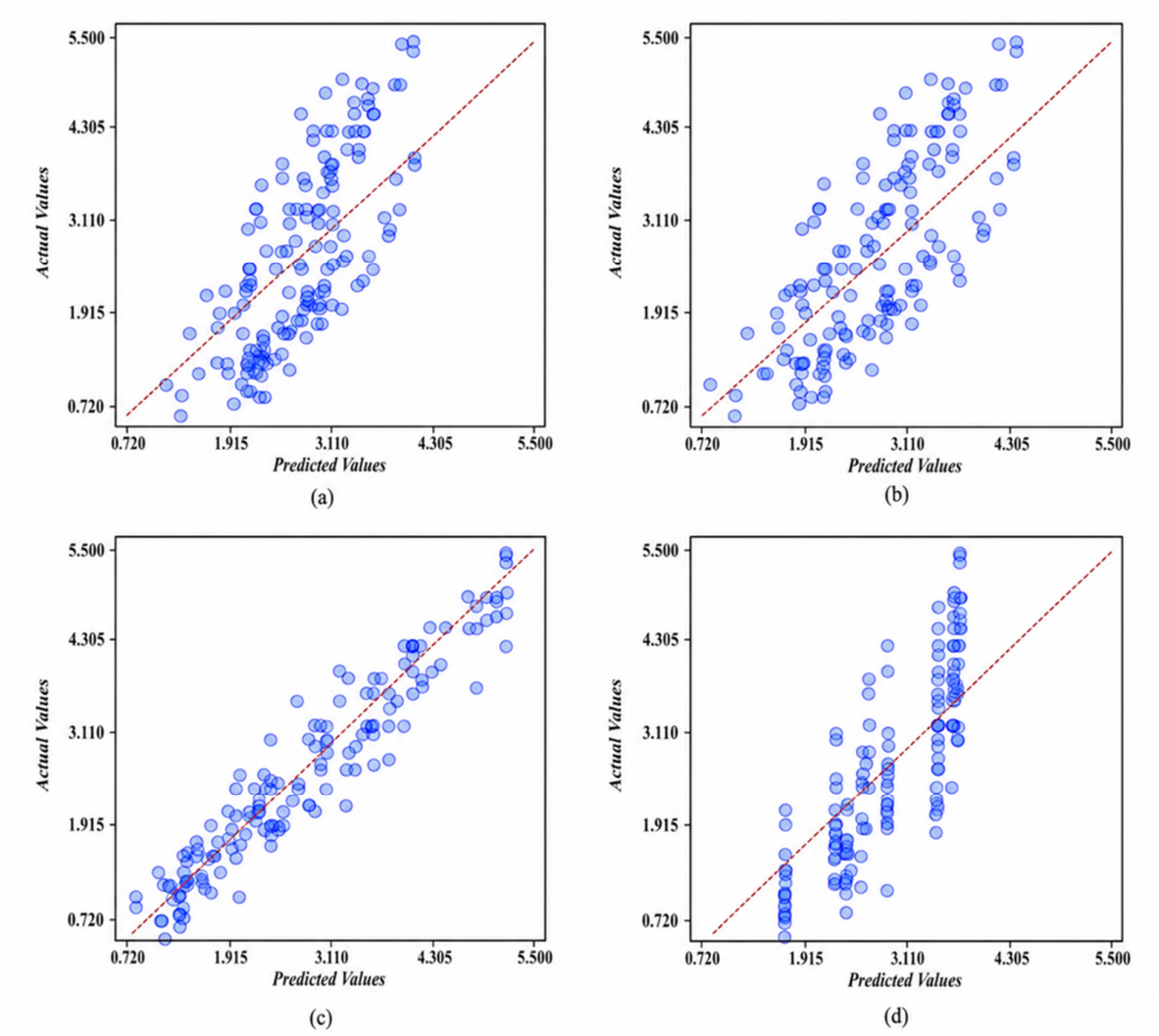

Figure 17. Comparison of predicted values versus observed data for the models: (a) Elastic Net, (b) Huber, (c) Ensemble, (d) Decision Tree

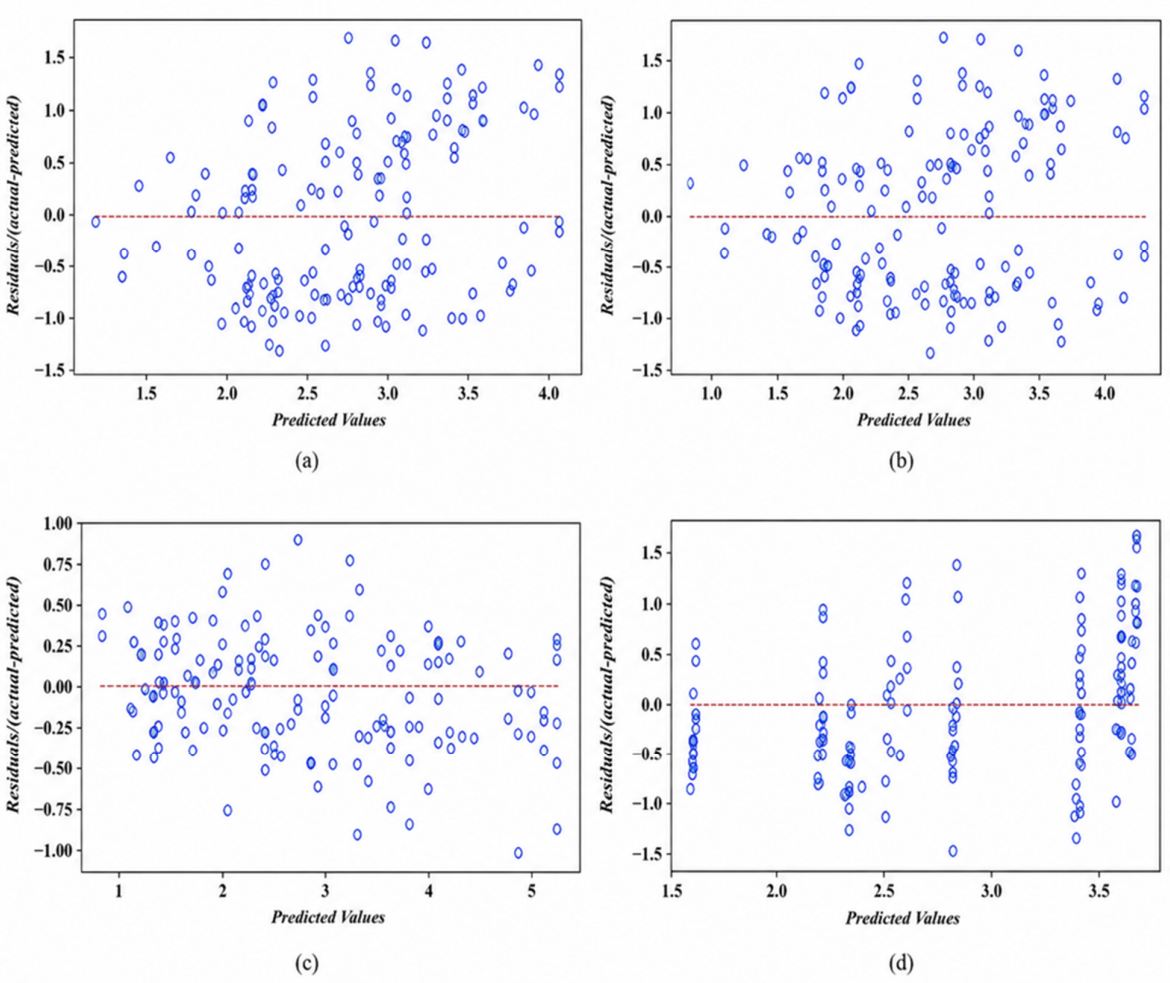

Figure 18. Diagnostic plots of model performance based on residuals versus fitted values: (a) Elastic Net, (b) Huber, (c) Ensemble, (d) Decision Tree

In predictive modeling based on input data and parameters, not all input variables typically contribute equally to variations in the output parameter. In other words, the dependency of the output parameter on each input variable is not uniform. Under such circumstances, understanding the influence of each input parameter on the model can be highly important. This process is commonly referred to as sensitivity analysis of the model with respect to its input parameters. By determining the degree of influence each input variable has on the output, valuable information can be obtained regarding the primary sources of error in predicting the output parameter, which can subsequently be used to control and manage model errors and behavior. Therefore, in this section of the study, a sensitivity analysis of the data has been conducted. In this research, the technique employed for sensitivity analysis is the relevancy factor. The relevancy factor is calculated using Equation (14).

$$rf(p_l, MMP) = \frac{\sum_{i=1}^{N} \left(p_{l,i} - \bar{p}_l\right)(MMP_i - \overline{MMP})}{\sqrt{\sum_{i=1}^{N}\left(p_{l,i} - \bar{p}_l\right)^2 \sum_{i=1}^{N}(MMP_i - \overline{MMP})^2}} \quad (14)$$

Where are:

rf – Relevancy Factor

$p_{l,i}$ – i -th value of l-th input parameter

$\bar{p}_l$ – Average of l-th input parameter

$\overline{MMP}$ – Average of predicted value

MMP_i – i-th value of predicted data

Using Equation (10) to calculate the relevancy factor for each input variable, the dependency of

the output parameter, penetration rate, on the input variables was determined. Figure 19 illustrates the relative influence of each input parameter on the behavior of the penetration rate. According to Figure 19, weight on bit has the highest impact, with a weighted influence of 50.83%. The subsequent rankings are the Schimazek's abrasivity index, bit rotation speed, and drilling fluid electrical conductivity, with respective weighted contributions of 27.95%, 15.31%, and 5.9%.

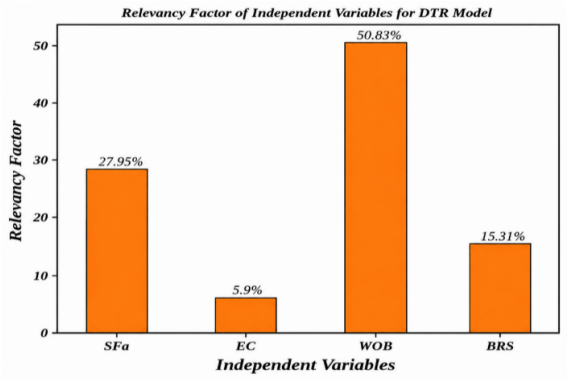

Figure 19. Influence of the weight of each input variable on the behavior of the penetration rate prediction model

The results of the sensitivity analysis indicated that the weight on bit (WOB) was the most influential parameter affecting the rate of penetration (ROP), accounting for 50.83% of the total contribution. This finding can be explained by the direct role of WOB in the rock breakage process. Increasing WOB increases the contact stress at the bit–rock interface and facilitates the initiation and propagation of tensile and shear fractures within the rock. As a result, rock fragmentation occurs more effectively and the penetration rate increases. In contrast, parameters such as bit rotation speed (BRS) and drilling fluid electrical conductivity (EC) mainly influence drilling conditions through their effects on cutting rate, lubrication, and heat dissipation. Therefore, their contribution to ROP is lower than that of WOB.

The obtained results also showed that the Decision Tree model outperformed the Huber, Elastic Net, and Ensemble models in predicting ROP. One possible explanation for this superior performance is the ability of the Decision Tree algorithm to capture complex and nonlinear relationships between the input variables and penetration rate. The drilling process is influenced by the interaction of several operational parameters and rock properties, and these relationships do not always follow linear trends. By recursively partitioning the data space into smaller regions, the Decision Tree model can better identify local patterns and hidden relationships within the dataset. Consequently, this model achieved higher prediction accuracy and lower error values compared with the other investigated methods.

It should be noted that the database used in this study was generated under controlled laboratory conditions and included only seven rock types. Therefore, the performance of the developed models may be affected when applied to geological and operational conditions different from those considered in the present study. In addition, factors such as geological heterogeneity, bit wear, drilling equipment vibrations, and operational uncertainties may influence drilling performance under actual mining conditions. Therefore, the use of field data obtained from real drilling operations in different mining environments is recommended to further evaluate the generalization capability of the developed models. Future studies should also incorporate larger and more diverse datasets, including different rock types and operating conditions, to further assess the robustness and applicability of the proposed models.

From a practical point of view, the results of this study suggest that the Decision Tree model can be used as a useful tool for predicting penetration rate and assisting drilling engineers in selecting appropriate operational parameters. Since all input variables used in the model can be measured or estimated during drilling operations, the proposed model may contribute to drilling optimization, improved operational efficiency, and reduced drilling costs. Nevertheless, further field-scale investigations are recommended before large-scale industrial implementation.

4. Conclusion

In this study, laboratory tests were conducted on 492 samples to investigate the effects of Schimazek's abrasivity index, drilling fluid electrical conductivity, weight on bit, and bit rotation speed on the rate of penetration (ROP). Using these parameters as input variables and ROP as the output variable, four machine learning techniques, including Huber Regression, Elastic Net, Ensemble, and Decision Tree, were developed and evaluated. The performance of the models was assessed using MAPE, R^2, RMSE, and EVS.

In general, the objective function values for the Huber, Elastic Net, Ensemble, and Decision Tree models were 1.2702, 1.4471, 0.5655, and 0.2975,

respectively. Based on these values, the Decision Tree model, which obtained the lowest objective function value, exhibited the best overall performance, while the Ensemble, Elastic Net, and Huber models ranked sequentially lower.

For qualitative performance assessment, several graphical analyses were conducted, including plotting predicted versus actual values against the data index, comparing predicted values against actual values, and visualizing predictions based on their residual errors. All of these analyses confirmed that the Decision Tree model provided the most accurate and reliable predictions. The superior performance of the Decision Tree model can be attributed to its ability to capture nonlinear relationships between the input variables and the rate of penetration, allowing it to better represent the underlying patterns within the dataset and achieve higher prediction accuracy than the other investigated methods.

From a practical point of view, the developed Decision Tree model shows considerable potential for application in real mining drilling operations. The model can be used as a tool for predicting the rate of penetration and optimizing drilling parameters. The implementation of this model may contribute to improving drilling efficiency, reducing operational costs, and enhancing the overall performance of drilling systems. However, further validation using field data is required before large-scale industrial application.

Despite the satisfactory results obtained in this study, several limitations should be acknowledged. The comparative analysis was limited to four machine learning algorithms, and the inclusion of additional methods such as Random Forest, XGBoost, LightGBM, Support Vector Regression (SVR), and Artificial Neural Networks (ANNs) may provide a more comprehensive evaluation of model performance. Furthermore, the hyperparameter tuning process was performed using a trial-and-error approach, which represents a methodological limitation of the present study and may introduce subjectivity into model optimization. Future studies should employ systematic hyperparameter optimization techniques to overcome this limitation and improve model calibration. Advanced interpretability methods such as SHAP analysis, the use of more diverse datasets, the investigation of additional influential parameters, and the application of statistical significance tests and confidence intervals may also contribute to the development of more accurate and reliable predictive models in future studies.

While a robust cross-validation framework is utilized in the current study to ensure model reliability and mitigate overfitting, it should be acknowledged that the hyperparameter tuning procedure was based on a trial-and-error approach. It is recognized that this manual strategy lacks the rigor of systematic optimization techniques such as Grid Search, Random Search, or Bayesian Optimization. Consequently, this is identified as a methodological limitation of the present work. It is suggested that future research would benefit from the implementation of these automated, systematic optimization frameworks to potentially further enhance predictive performance and model generalizability.

Conflicts of interest

The authors declare no conflict of interest.

Funding Declaration

This research received no funding.

Data availability

All data generated and analyzed during this study are included in this article.

References

[1] Kolapo, P. (2021). Investigating the effects of mechanical properties of rocks on specific energy and penetration rate of borehole drilling. *Geotechnical and Geological Engineering*, *39*(2), 1715-1726.

[2] Komadja, G. C., Westman, E., Rana, A., & Vitalis, A. (2025). Predicting rock mass strength from drilling data using synergistic unsupervised and supervised machine learning approaches. *Earth Science Informatics*, *18*(3), 325.

[3] Ataei, M., KaKaie, R., Ghavidel, M., & 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.

[4] Hosseini, S. H., Ataie, M., & Aghababaie, H. (2014). A laboratory study of rock properties affecting the penetration rate of pneumatic top hammer drills. *Journal of mining and environment*, *5*(1), 25-34.

[5] Kahraman, S. A. I. R., Bilgin, N., & Feridunoglu, C. (2003). Dominant rock properties affecting the penetration rate of percussive drills. *International Journal of Rock Mechanics and Mining Sciences*, *40*(5), 711-723.

[6] Khosravimanesh, S., Mikaeil, R., Esmaeilzadeh, A., Piri, M., & Mirahmadi, M. (2026). Mitigating Drill Bit Vibration through Cooling-Lubrication Fluids: An Experimental Study on the Interplay of Fluid Properties

and Rock Mechanics. *Journal of Vibration Engineering & Technologies*, *14*(7), 451.

[7] Amadi, K., & Iyalla, I. (2012, June). Application of mechanical specific energy techniques in reducing drilling cost in deepwater development. In *SPE deepwater drilling and completions conference* (pp. SPE-156370). SPE.

[8] Khosravimanesh, S., Cheraghi Seifabad, M., Mikaeil, R., & Bagherpour, R. (2024). Studying Effects of Cooling/Lubricating Fluids, Machining Parameters, and Rock Mechanical Properties on Specific Energy in Rock Drilling Process. *Journal of Mining and Environment*, *15*(2), 613-636.

[9] Suppes, R., Ebrahimi, A., & Krampe, J. (2019). Optimising casing milling Rate of Penetration (ROP) by applying the concept of Mechanical Specific Energy (MSE): a justification of the concept's applicability by literature review and a pilot study. *Journal of Petroleum Science and Engineering*, *180*, 918-931.

[10] Detournay, E., & Defourny, P. (1992, January). A phenomenological model for the drilling action of drag bits. In *International journal of rock mechanics and mining sciences & geomechanics abstracts*, *29*(1), 13-23.

[11] Zhou, Y., Zhang, W., Gamwo, I., & Lin, J. S. (2017). Mechanical specific energy versus depth of cut in rock cutting and drilling. *International Journal of Rock Mechanics and Mining Sciences*, *100*, 287-297.

[12] Chen, X., Fan, H., Guo, B., Gao, D., Wei, H., & Ye, Z. (2014). Real-time prediction and optimization of drilling performance based on a new mechanical specific energy model. *Arabian Journal for Science and Engineering*, *39*(11), 8221-8231.

[13] El-Biblawi, M. M., Sayed, M. A., Mohamed, M. T., & El-Rawy, W. R. (2007). Some drilling parameters as a tool to predict different categories of rocks. *JES. Journal of Engineering Sciences*, *35*(4), 995-1008.

[14] Saffari, A., Ataei, M., & Sereshki, F. (2020). Studying Relationship between Coal Intrinsic Characteristics in Spontaneous Combustion of Coal Potential Using Crossing Point Temperature Test Method. *Journal of Mining and Environment, 11*(1), 315–333.

[15] Anemangely, M., Ramezanzadeh, A., & Tokhmechi, B. (2017). Determination of constant coefficients of Bourgoyne and Young drilling rate model using a novel evolutionary algorithm. *Journal of Mining and Environment*, *8*(4), 693-702. [16] Yarali, O., & Soyer, E. (2013). Assessment of relationships between drilling rate index and mechanical properties of rocks. *Tunnelling and underground space technology*, *33*, 46-53.

[17] Taheri, A., Qao, Q., & Chanda, E. (2016). Drilling penetration rate estimation using rock drillability characterization index. *Journal of The Institution of Engineers (India): Series D*, *97*(2), 159-170.

[18] Tiryaki, B. (2008). Predicting intact rock strength for mechanical excavation using multivariate statistics, artificial neural networks, and regression trees. *Engineering Geology*, *99*(1-2), 51-60.

[19] Shad, H. I. A., Sereshki, F., Ataei, M., & Karamoozian, M. (2018). Prediction of rotary drilling penetration rate in iron ore oxides using rock engineering system. *International Journal of Mining Science and Technology*, *28*(3), 407-413. [20] Khosravimanesh, S., Seifabad, M. C., Mikaeil, R., & Bagherpour, R. (2021). Experimental study of the effect of cooling lubricating fluids on penetration rate in a hard and soft rock drilling process. *Rudarsko-geološko-naftni zbornik*, *36*(5), 79-91.

[21] Khosravimanesh, S., Cheraghi Seifabad, M., Mikaeil, R., & Bagherpour, R. (2021). Study of effect of cooling/lubricating fluids, machining parameters, and rock mechanical properties on penetration rate in rock drilling process. *Journal of Mining and Environment*, *12*(3), 825-843.

[22] Khosravimanesh, S., Esmaeilzadeh, A., Akhyani, M., Mikaeil, R., & Asl, M. M. (2024). Accurate prediction of drill bit penetration rate in rock using supervised machine learning techniques base on laboratory test data. *Rudarsko-geološko-naftni zbornik*, *39*(1), 115-130.

[23] Singh, T. N., Jain, A., & Sarkar, K. (2009). Petrophysical parameters affecting the microbit drillability of rock. *International Journal of Mining and Mineral Engineering*, *1*(3), 261-277.

[24] Heydari, S., Hoseinie, S. H., & Bagherpour, R. (2024). Prediction of jumbo drill penetration rate in underground mines using various machine learning approaches and traditional models. *Scientific Reports*, *14*(1), 8928.

[25] Yenice, H. (2019). Determination of drilling rate index based on rock strength using regression analysis. *Anais da Academia Brasileira de Ciências*, *91*(03), e20181095.

[26] Yenice, H., Özdoğan, M. V., & Özfirat, M. K. (2018). A sampling study on rock properties affecting drilling rate index (DRI). *Journal of African Earth Sciences*, *141*, 1-8.

[27] Saksala, T., Fourmeau, M., Kane, P. A., & Hokka, M. (2018). 3D finite elements modelling of percussive rock drilling: estimation of rate of penetration based on multiple impact simulations with a commercial drill bit. *Computers and Geotechnics*, *99*, 55-63.

[28] Sazidy, M. S., Rideout, D. G., Butt, S. D., & Arvani, F. (2010, June). Modeling percussive drilling performance using simulated visco-elasto-plastic rock medium. In *ARMA US Rock Mechanics/Geomechanics Symposium* (pp. ARMA-10).

[29] Rossi, E., Kant, M., Borkeloh, O., Saar, M. O., & Rudolf von Rohr, P. (2019). Experiments on rock-bit interaction during a combined thermo-mechanical drilling method. In *43rd Workshop on Geothermal Reservoir Engineering 2018. Proceedings of a meeting held 12-14 February 2018, Stanford, California, USA*, *1*, 1874.

[30] Rossi, E., Saar, M. O., & Rudolf von Rohr, P. (2020). The influence of thermal treatment on rock–bit interaction: a study of a combined thermo–mechanical drilling (CTMD) concept. *Geothermal Energy*, *8*(1), 16.

[31] Zhong, J., Yang, J., & Butt, S. (2016, April). DEM simulation of enhancing drilling penetration using vibration and experimental validation. In *SpringSim (ANSS)* (p. 12).

[32] Dong, G., & Chen, P. (2018). 3D numerical simulation and experiment validation of dynamic damage characteristics of anisotropic shale for percussive-rotary drilling with a full-scale PDC bit. *Energies*, *11*(6), 1326.

[33] Zhang, H., Ni, H., Huang, B., Liu, S., Wang, Y., Liang, H., & Guo, X. (2022). Research on discrete element modeling and numerical simulation of cutting rock behavior under impact load. *Energy Science & Engineering*, *10*(7), 2420-2436.

[34] Zhang, C., Wang, Y., Xuan, L., Ren, H., Yang, Y., & Sun, M. (2023). Numerical simulation and experimental study on the interaction between a convex ring PDC bit and rock. *Geoenergy Science and Engineering*, *230*, 212247.

[35] Ma, Y., Gong, Q., Zhou, X., Yin, L., & Ma, H. (2023). The modelling of rock fragmentation mechanisms by carbide buttons using the 3D discrete element method. *Applied Sciences*, *13*(10), 6090.

[36] Li, S., Wang, S., Li, Z., Zhou, Y., Gao, S., & Li, W. (2025). Experiment and numerical simulation of rock breaking by micro PDC drill bit under harmonic dynamic loading. *Rock Mechanics and Rock Engineering*, *58*(12), 13809-13826.

[37] Yang, Y., Cen, X., Ni, H., Liu, Y., Chen, Z. J., Yang, J., & Hong, B. (2025). A highly accurate and robust prediction framework for drilling rate of penetration based on machine learning ensemble algorithm. *Geoenergy Science and Engineering*, *244*, 213423.

[38] Barbosa, L. F. F., Nascimento, A., Mathias, M. H., & de Carvalho Jr, J. A. (2019). Machine learning methods applied to drilling rate of penetration prediction and optimization-A review. *Journal of Petroleum Science and Engineering*, *183*, 106332.

[39] Kahraman, S. (2016). The prediction of penetration rate for percussive drills from indirect tests using artificial neural networks. *Journal of the Southern African Institute of Mining and Metallurgy*, *116*(8), 793-800.

[40] Azar, H. F., Saksala, T., & Jalali, S. M. E. (2017). Artificial neural networks models for rate of penetration prediction in rock drilling. *Rakenteiden mekaniikka*, *50*(3), 252-255.

[41] Basarir, H., Tutluoglu, L., & Karpuz, C. (2014). Penetration rate prediction for diamond bit drilling by adaptive neuro-fuzzy inference system and multiple regressions. *Engineering Geology*, *173*, 1-9.

[42] Gan, C., Cao, W. H., Wu, M., Chen, X., Hu, Y. L., Liu, K. Z., ... & Zhang, S. B. (2019). Prediction of drilling rate of penetration (ROP) using hybrid support vector regression: A case study on the Shennongjia area, Central China. *Journal of Petroleum Science and Engineering*, *181*, 106200.

[43] Kazemi, M. M. K., Nabavi, Z., & Armaghani, D. J. (2024). A novel hybrid XGBoost methodology in predicting penetration rate of rotary based on rock-mass and material properties. *Arabian Journal for Science and Engineering*, *49*(4), 5225-5241.

[44] Zhou, F., Fan, H., Liu, Y., Ye, Y., Diao, H., Wang, Z., ... & Davio, E. (2022, February). Application of xgboost algorithm in rate of penetration prediction with accuracy. In *International Petroleum Technology Conference* (p. D012S111R002). IPTC.

[45] Liu, N., Gao, H., Zhao, Z., Hu, Y., & Duan, L. (2022). A stacked generalization ensemble model for optimization and prediction of the gas well rate of penetration: a case study in Xinjiang. *Journal of Petroleum Exploration and Production Technology*, *12*(6), 1595-1608.

[46] Yang, X., Xiang, J., Naderi, S., Wang, Y., Aising, J., Ugarte, I., & Latham, J. P. (2025). Multi-criteria validation of hi-fidelity numerical model of impact breakage: towards next generation percussion drill simulation. *International Journal of Rock Mechanics and Mining Sciences*, *191*, 106125.

[47] Shahani, N. M., Zheng, X., Wei, X., & Hongwei, J. (2024). Hybrid machine learning approach for accurate prediction of the drilling rate index. *Scientific reports*, *14*(1), 24080.

[48] Mohammadinia, F., Ranjbar, A., Ghazi, F., & Hosseini, S. T. (2025). Rate of penetration prediction in drilling operations: a comparative study of AI models and meta-heuristic approaches. *Journal of Petroleum Exploration and Production Technology*, *15*(7), 115.

[49] Esmaeilzadeh, A., Mikaeil, R., Khosravimanesh, S., Shaffiee Haghshenas, S., Simic, V., & Pamucar, D. (2025). Biogeography-Based Optimization of Machine Learning Models for Accurate Penetration Rate Prediction Using Rock Texture Coefficient. *International Journal of Computational Intelligence Systems*, *18*(1), 226.

[50] Khamis, Y. E., El-Rammah, S. G., & Salem, A. M. (2023). Rate of penetration prediction in drilling operation in oil and gas wells by k-nearest neighbors and

multi-layer perceptron algorithms. *Journal of Mining and Environment*, *14*(3), 755-770.

[51] Mikaeil, R., Piri, M., Shaffiee Haghshenas, S., Careddu, N., & Hashemolhosseini, H. (2022). An experimental-intelligent method to predict noise value of drilling in dimension stone industry. *Journal of Mining and Environment*, *13*(3), 693-713.

[52] Huang, G., Hamzah, S. K., Patel, P., Ramachandran, T., Shankhyan, A., Karthikeyan, A., ... & Abbasi, H. (2025). Data-driven prediction of rate of penetration (ROP) in drilling operations using advanced machine learning models. *Journal of Petroleum Exploration and Production Technology*, *15*(7), 121.

[53] Oyedere, M., & Gray, K. (2020). ROP and TOB optimization using machine learning classification algorithms. *Journal of Natural Gas Science and Engineering*, *77*, 103230.

[54] Eskandarian, S., Bahrami, P., & Kazemi, P. (2017). 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*, 605-615.

[55] Elkatatny, S., Al-AbdulJabbar, A., & Abdelgawad, K. (2020). A new model for predicting rate of penetration using an artificial neural network. *Sensors*, *20*(7), 2058.

[56] Riazi, M., Mehrjoo, H., Nakhaei, R., Jalalifar, H., Shateri, M., Riazi, M., ... & Hemmati-Sarapardeh, A. (2022). Modelling rate of penetration in drilling operations using RBF, MLP, LSSVM, and DT models. *Scientific Reports*, *12*(1), 11650.

[57] Ben Aoun, M. A., & Madarász, T. (2022). Applying machine learning to predict the rate of penetration for geothermal drilling located in the Utah FORGE site. *Energies*, *15*(12), 4288.

[58] Ren, C., Huang, W., & Gao, D. (2023). Predicting rate of penetration of horizontal drilling by combining physical model with machine learning method in the China Jimusar Oil Field. *SPE Journal*, *28*(06), 2713-2736.

[59] Junior, J. R. B., & Lavi, B. (2025). Enhancing rate of penetration prediction in drilling operations: A data stream framework approach. *Engineering Applications of Artificial Intelligence*, *143*, 110034.

[60] Plinninger, R. J. (2008). Abrasiveness assessment for hard rock drilling. *Geomechanik und Tunnelbau: Geomechanik und Tunnelbau*, *1*(1), 38-46.

[61] Schimazek, J., & Knatz, H. (1970). Der Einfluß des Gesteinsaufbaus auf die Schnittgeschwindigkeit und den Meißelverschleiß von Streckenvortriebsmaschinen. *Glückauf*, *106*(6), 274-278.

[62] Goktan RM & Gunes Yilmaz, N. (2005). A new methodology for the analysis of the relationship between rock brittleness index and drag pick cutting efficiency. *Journal of the Southern African Institute of Mining and Metallurgy*, *105*(10), 727-733.

[63] Sirdesai, N., Aravind, A., & Singh, A. (2019, December). Correlation of Abrasivity and Physico-mechanical properties of rocks: An Experimental, Statistical and Soft-computing analysis. In *ISRM Young Scholars Symposium on Rock Mechanics* (pp. ISRM-YSRM). ISRM.

[64] Fairhurst, C. E., & Hudson, J. A. (1999). Draft ISRM suggested method for the complete stress-strain curve for intact rock in uniaxial compression. *International journal of rock mechanics and mining sciences (1997)*, *36*(3), 279-289.

[65] Huber, P. J. (1992). Robust estimation of a location parameter. In *Breakthroughs in statistics: Methodology and distribution* (pp. 492-518). New York, NY: Springer New York.

[66] Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *67*(2), 301-320.

[67] Zhou, Z. H. (2025). *Ensemble methods: foundations and algorithms*. CRC press.

[68] Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, *1*(1), 81-106.

نشریه مهندسی معدن و محیط زیست
نشانی نشریه: www.jme.shahroodut.ac.ir
انجمن مهندسی معدن ایران
دانشگاه صنعتی شاهرود

بهره‌گیری از شاخص سایندگی شیمازک و هدایت الکتریکی سیال حفاری برای بهبود پیش‌بینی نرخ نفوذ حفاری: یک مطالعه مقایسه‌ای یادگیری ماشین

شاهرخ خسروی‌منش[۱]، اکبر اسمعیل‌زاده[۲]، رضا میکائیل[۲*] و ساسان قربانی[۳]

۱. گروه مهندسی معدن، دانشگاه صنعتی اصفهان (IUT)، اصفهان، ایران
۲. گروه مهندسی معدن، دانشکده محیط زیست، دانشگاه صنعتی ارومیه(UUT) ، ارومیه، ایران
۳. گروه مهندسی معدن، دانشگاه صنعتی همدان، همدان، ایران

اطلاعات مقاله

تاریخ ارسال: ۲۰۲۶/۰۲/۲۳
تاریخ داوری: ۲۰۲۶/۰۷/۱۸
تاریخ پذیرش: ۲۰۲۶/۰۸/۳۱

DOI: 10.22044/jme.2026.17539.3489

کلمات کلیدی

نرخ نفوذ
هدایت الکتریکی
رگرسیون هابر
شبکه الاستیک
درخت تصمیم

چکیده

نرخ نفوذ حفاری (ROP) یکی از شاخص‌های کلیدی در عملیات حفاری است که به‌طور مستقیم بر برآورد هزینه و زمان‌بندی پروژه تأثیر می‌گذارد. اندازه‌گیری مستقیم ROP معمولاً پرهزینه بوده و از نظر کاربرد محدود است، بنابراین توسعه مدل‌های پیش‌بینی دقیق مبتنی بر داده از اهمیت زیادی برخوردار است. در این مطالعه، مجموعه‌داده‌ای شامل ۴۹۲ اندازه‌گیری آزمایشگاهی گردآوری شد که چهار متغیر مستقل شامل شاخص سایندگی شیمازک(SFa) ، هدایت الکتریکی سیال حفاری(EC) ، وزن روی مته (WOB) و سرعت چرخش مته (BRS) را دربر می‌گرفت و ROP به عنوان متغیر وابسته در نظر گرفته شد. چهار الگوریتم یادگیری ماشین نظارت‌شده رگرسیون هابر، شبکه الاستیک، انسمبل و درخت تصمیم برای پیش‌بینی ROP به کار گرفته شدند. عملکرد پیش‌بینی مدل‌ها با استفاده از معیارهای MAPE، R^2، RMSE و EVS ارزیابی شد. علاوه بر این، یک تابع هدف ترکیبی با تلفیق این شاخص‌ها تعریف شد تا رتبه‌بندی کلی مدل‌ها فراهم شود. نتایج نشان داد که مدل درخت تصمیم با مقدار ۰.۲۹۷۵ برای تابع هدف، بهترین عملکرد را به دست آورد و پس از آن، مدل‌های انسمبل، شبکه الاستیک و رگرسیون هابر به‌ترتیب با مقادیر ۰.۵۶۵۵، ۱.۴۴۷۱ و ۱.۲۷۰۲ قرار گرفتند. این یافته‌ها نشان می‌دهند که الگوریتم درخت تصمیم دقیق‌ترین و قابل‌اعتمادترین پیش‌بینی‌ها را برای ROP ارائه می‌دهد و می‌تواند به‌عنوان ابزاری مؤثر برای بهینه‌سازی زمان حفاری و کاهش هزینه‌های عملیاتی مورد استفاده قرار گیرد.

نویسنده مسئول مکاتبات: *reza.mikaeil@gmail.com* (ر. میکائیل)