Shahrood University of Technology

Iranian Society of Mining Engineering (IRSME)

Journal of Mining and Environment (JME)

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

THOMSON REUTERS

A Probabilistic Approach for Prediction of Drilling Rate Index using Ensemble Learning Technique

Muhammad Kamran*

Department of Mining Engineering, Bandung Institute of Technology, Kota Bandung, Indonesia

Article Info

Received 7 April 2021

Received in Revised form 27 April 2021

Accepted 4 May 2021

Published online 4 May 2021

DOI:10.22044/jme.2021.10689.2030

Keywords

Drilling rate index

Ensemble learning

Prediction

Drillability

Probability

Abstract

Drillability is one of the significant issues in rock engineering. The drilling rate index (DRI) is an important tool in analyzing the drillability of rocks. Several efforts have been made by the researchers to correlate and evaluate DRI of rocks. The ensemble learning methods including the decision tree (DT), adaptive boosting (AdaBoost), and random forest (RF) are employed in this research work in order to predict DRI of rocks. A drillability database with four parameters is compiled in this work. A relationship between the input parameters and DRI is established using the simple regression analysis. In order to train the model, different mechanical properties of rocks incorporating the uniaxial compressive strength (UCS), Brazilian tensile strength (BTS), brittleness test (S20), and sievers' J-miniature drill value (Sj) are taken as the input variables. The original DRI database is randomly divided into the training and test sets with an 80/20 sampling method. Various algorithms are developed, and consequently, several approaches are followed in order to predict DRI of the rock samples. The model performance has revealed that RF predicts DRI with a high accuracy rate. Besides, the Monte Carlo simulations exhibit that this approach is more reliable in predicting the probability distribution of DRI. Therefore, the proposed model can be practiced for the stability risk management and the investigative design of DRI.

1. Introduction

Drillability of rocks is defined as the speed at which a drill bit penetrates a rock mass. The drilling rate index (DRI) is an important tool in approximating the drillability of various rock samples [1]. DRI is influenced by the multiple uncertain attributes, which are classified into the controllable and uncontrollable attributes [2]. The rotation speed, pumping rate, weight-on-bit, torque, and standpipe pressure are considered as the controllable attributes. In contrast, the drill bit size, fluid nature, density, and physico-chemical properties are considered as the uncontrollable attributes [3-5]. The deep drillability analysis can assist in providing a good description of the rock attributes encountered during a field [6]. The fluid properties and reservoir features have been examined from several rock specimens from target formation during the well performance optimization [7]. The structural parameters of rock mass play an important role in the evaluation of drilling rate [8]. DRI greatly influences the mine project design and budget of the mining industry [9]. The impact of rock hardness on the penetration rate of pneumatics percussive drills in the experimental laboratory has been utilized using the Mohs hardness, indentation hardness index, and L-type Schmidt hammer [10]. The researchers have made several efforts to evaluate and correlate DRI of rocks [11-14]. The empirical methods are not an

Corresponding author: muhdkamran62@gmail.com (M. Kamran).

excellent approach to handle the non-linear and multivariable equations [15].

Most recently, the application of machine learning techniques has been utilized to predict the drillability of rocks. A DRI evaluation model has been suggested based on the fuzzy Delphi analytic hierarchy method. Based on the drillability classification, the rocks are labeled into five groups varying from very poor to very good [16]. A Fuzzy system computing approach has been proposed that predicts the drilling rate with an acceptable range of accuracy [17-19]. A novel artificial neural network (ANN) model shows that it can be executed well in predicting DRI examined for auxiliary renovated frameworks [20-24]. A regression model based on weak learners has been proposed using the improved machine learning algorithm, enhancing the model's prediction performance [25-26]. The advancement of intelligent predictive models has made it possible to select the drill bit. Different machine learning algorithms, e.g. support vector machine, K-nearest neighbors, naïve Bayes, and multilayer perceptron have been assessed to understand the allowable declining causes of the disproportion drilling data [27]. In order to reduce the drilling engineering costs, the interaction of machine learning algorithms and drilling engineering technology may allow up-to-date techniques for raising the rate of penetration (ROP) [28]. The Artificial Bee Colony (ABC) algorithm has been utilized to systemize ROP by employing a unique robust high-level meta-heuristic algorithm and self-organizing map as one of the authentic technological mechanisms [29]. The diamond bit drilling has been extensively operated in different rock engineering operations considering high ROP, core recovery, and its capacity to penetrate the rock mass with a minor variance deviation. The establishment of multiple regression models and an adaptive neuro-fuzzy inference system for forecasting ROP of diamond drilling has been suggested [30]. Monte Carlo simulation is a suitable method for modeling and evaluating the irregularity in the penetration parameters of rock mass [31].

However, the premonitory characteristics for the DRI prediction are not perpetually consistent in numerous geologic conditions. Although various promising results have been achieved in the multiple aspects of the DRI analysis, the probabilistic DRI prediction remains problematic. Presently, there is no steady approach in the engineering practice.

In this research work, a comparative analysis of ensemble learning is proposed for predicting DRI. The ensemble learning is a machine learning approach that consists of different models in order to realize the problem where a discrete model inaccuracy is possibly re-compensated by the other models [32]. The standard ensemble learning methods consist of the regression tree [33], random forest (RF) [34], and adaptive boosting (AdaBoost) [35]. The Monte Carlo (MC) simulation method has been proposed for the direct calculation of the expanded ensemble [36-37]. To the best of our knowledge, this is the first work to predict DRI by the comparative analysis of the ensemble learning algorithms.

The work`s primary objective is to develop a probabilistic model for DRI prediction based on the ensemble learning method. Firstly, the mechanical properties of the rock samples are extracted from the published literature. Secondly, three ensemble learning algorithms are employed in order to predict the model. Lastly, the comparative predictive performance indicator of every model is thoroughly examined.

Moreover, in this research work, the author discusses four aspects of DRI, summarized as follow:

i) Statistical analysis is performed on the DRI database in order to obtain their average, standard deviation, and minimum and maximum values.

ii) Establishing the relationships between the physical and mechanical properties of DRI.

iii) Construction of various ensemble learning approaches by checking its performance using the input and output datasets.

iv) Selection of the best model for the DRI prediction.

v) Monte Carlo simulation in order to predict the range of uncertainty in the DRI phenomenon.

2. Methodology

2.1. Set range of input database

A quick and convenient approach to a large number of output values is the inferiority of artificial intelligence. An excellent database always requires a high-quality data closely resembling the real-world problems. Several types of rocks collected from the standard laboratory samples ranging from weak to very strong rocks including the metamorphic, magmatic, and sedimentary types were compiled in this work. From the previous literature, 57 datasets [38-39] were collected to predict DRI of various rock samples, as shown in appendix 1. There are four specific input attributes, i.e. the Uniaxial Compressive Strength (UCS), Brazilian Tensile Strength (BTS), Brittleness test (S_{20}), and Sievers' J-miniature drill value (S_j). Additionally, DRI is taken as an output parameter. The descriptive statistics of the mechanical properties for the rock samples datasets analyzed during the work are outlined in Table 1.

Table 1. Descriptive statistics of mechanical properties of rock sample datasets.

	UCS (MPa)	BTS (MPa)	Brittleness value (S_{20})	Sievers' J value (S_j)	DRI
Average	102.6602	8.114035	51.38947	41.64807	56.91228
Standard deviation	47.10728	2.592484	12.02135	25.5928	12.65374
Maximum	206.4	13.82	82.83	87	89
Minimum	23.43	3.28	30.02	3.6	25

2.2 Application of simple regression analysis in DRI phenomenon

Simple Regression Analysis (SRA) is a tool to determine a relation between two or more independent indicators and one dependent indicator in statistics and machine learning. This approach is employed in numerous models in order to identify how well a set of variables explain a phenomenon [40]. SRA was employed in the drilling data, showing the highest coefficient of correlation [10]. This technique is widely used to approximate various issues associated with rock engineering by getting the best-fit mathematical equation. In order to predict DRI of rocks employing an SRA technique, the measured DRI values are calculated as the product of the four input indicators, namely UCS, BTS, (S_{20}), and (S_j). The developed simple regression equation for the prediction of DRI is shown in Equation 1.

$$\mathrm{DRI} = -0.013\ \mathrm{UCS} - 0.13\ \mathrm{BTS} + 0.97\ \mathrm{S_{20}} + 0.14\ \mathrm{S_J} + 3.19 \quad (1)$$

2.3. Overview of box plot from DRI database

Box plot is a graphical technique that is applied to correlate and summarize the data groups. The box plot utilizes the median, estimated quartiles, and smallest and highest data points to measure the magnitude and consistency of data dissemination. The data examined by a box plot mechanism has superiority over a data table. It is a technique that enhances and refines the reasoning behind the quantitative knowledge [41].

The box plot of each indicator for DRI is provided in Figure 1. DRI is positively correlated with UCS and S_{20}, and negatively correlated with BTS and S_j. The indicator values depend upon the DRI value. The higher the indicator values, the larger the DRI value. However, some outliers stay in all parameters under every indicator, which shows the entanglement of the DRI development mechanism. Furthermore, for the identical parameters, the distance between the larger and the smaller quartiles (box' height) varies. The alignment of the parameter values also has some overlapping segments. Hence, the impact of all parameters is assimilated to perceive a higher accuracy of DRI.

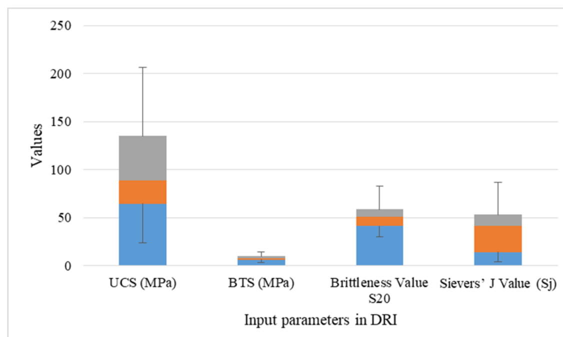

Figure 1. Input parameters of the box plot from DRI database.

3. A brief introduction of machine learning algorithms

This section concisely explains the theoretical framework of DT, AdaBoost, and RF employed in this work.

3.1. Decision tree modeling

A decision tree (DT) is among the most famous ensemble techniques that represent a systemic sample of features as the tree origination. DT assists in explaining the features and sorting out the predictions based on the features [42]. DT manipulates the data measured on several model distributions, build on a simple non-linear relationship [43-44]. The ensemble learning models typically require an excessive numeration. Consequently, the fast machine learning algorithms, namely DT, are accustomed to develop the ensemble models [45].

3.2. Adaboost algorithm

AdaBoost is one of the most major ensemble methods in machine learning. This algorithm was initially developed on the weak classifier models in order to reinforce their accuracy [46]. The Adaboost algorithm can be used for both the classification and regression problems [47]. The machine learning algorithm has its desirability resulting in more convenience for several types of models, and there are ordinarily various attributes, configurations, and parameters that are required to be changed before attaining the most compelling interpretation of a model [48].

3.3. Random forest approach

RF is a widely used ensemble learning to predict different categories based on various databases [49]. RF permits an unprecedented collaboration of model illustration and prediction accuracy among the alternative prominent machine learning algorithms. The ensemble learning used in RF allows it to attain a higher generalization and correct the model prediction [50]. Furthermore, RF focuses on three critical aspects, described as follow:

i) It provides decent information for prediction purposes.

ii) It can compute the significance of every parameter with the training of various models.

iii) Nodesets proximity among different samples can be calculated by the model training.

4. Model performance indicators

In order to accurately and effectively evaluate the DT, AdaBoost, and RF models, three different evaluation analysis indicators are employed to explain the correlation between the predictive values and the measured values, namely the Pearson correlation coefficient (PCC), root mean square error (RMSE), and mean absolute error (MAE). PCC is a quantity index used to calculate how strong the correlation is between two variables, as shown in Equation 2, RMSE is a customary applied statistical metric that shows the fitted standard deviation of the error between the predictive values and the actual value, given in Equation 3. MAE is another helpful measure widely employed in the model evaluations, showing the mean of the absolute error, which can well reflect the exact condition of the predictive value error. The MAE value can be calculated using Equation 4.

$$\mathrm{PCC}^2 = 1 - \frac{\sum_{i=1}^{n}(X - X')^2}{\sum_{i=1}^{n}(X - X")^2} \tag{2}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(X - X")^2} \tag{3}$$

$$MAE = \sum_{i=1}^{n}| X'' - X| \tag{4}$$

where, n is the total number of databases, X is the actual value, X' indicates the predicted value, and X'' represent the mean values.

5. Structure of work

In this approach, DT, AdaBoost, and RF are employed in order to predict DRI, and their prognostic interpretation is thoroughly compared from different aspects. First, the DRI database is randomly divided into a training dataset (80% of the total database) and the testing dataset (20% of the database). It should be noticed that the distribution of the DRI samples with various indictors in the training and test sets is maintained consistently during the database partitioning. Secondly, a three-fold cross-validation (CV) methodology is followed in order to get the best hyper-parameter values of the three ensemble learning algorithms (DT, AdaBoost, and RF). Thirdly, every algorithm with adjusted hyper-parameters is then determined by the performance of the training set. Fourthly, the test set is acquired in order to examine the model performance using the performance analysis indicators, namely, PCC, root mean square error (RMSE), and mean absolute error (MAE). Fifthly, after all, the models are evaluated, one of them is selected for deployment. Finally, the Monte Carlo simulation is applied in order to predict the best simulation range of the model. The detailed framework of the work is shown in Figure 2.

6. Ensemble learning

In keeping with the dependencies among the base learners, bagging is a significant type of ensemble learning method [51]. The base learners are usually weak, having no dependencies, which permits them to be implemented in a parallel pattern in bagging ensemble learning [52]. First, a bootstrap illustration method is modeled in order to create the trail sets from the preliminary database. Then, the base learners are independently trained using each trail set. Finally, the predicted value is achieved using the integration rules. A voting technique is formed for the prediction of the model. RF is the conventional representation of the bagging approach. However, in the selection of the features, RF varies from a bagging ensemble learning. The features are randomly taken from the original features within the RF, enhancing the model logical reasoning capacity. Figure 3. Indicates the structure of the bagging ensemble learning applied in the work.

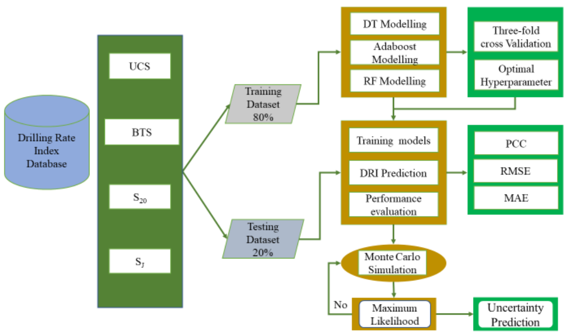

Figure 2. A flow chart of the work.

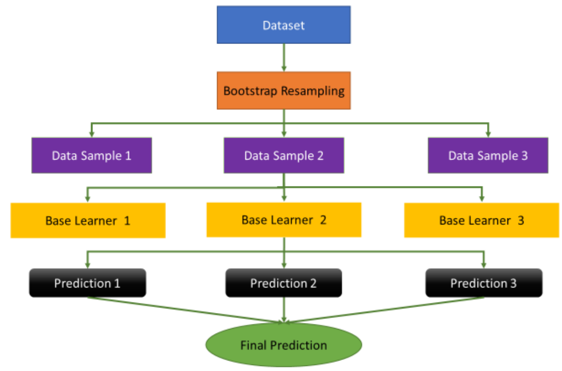

Figure 3. Diagram of bagging ensemble learning.

7. Discussion

Various architectures were employed in order to obtain the best prediction of DRI in this work. To demonstrate that the proposed prediction ensemble learning model has a greater predictive capacity, the predictive results of the three models are compared with each other.

Figures 4 and 5 indicate the predictive results of the training and testing datasets, respectively. It is clear that the PCC squared value of the RF model is near to 1 than that of the AdaBoost and DT models, which shows that the predictive values of DRI by the RF model are more correlated with the measured values. Simultaneously, the RMSE, and MAE values of the RF predictive model are lower than those of AdaBoost and DT, indicating that the RF model has a smaller prediction error. Hence it can be concluded that the performance rank of the ensemble learning in the descending order is RF > AdaBoost > DT.

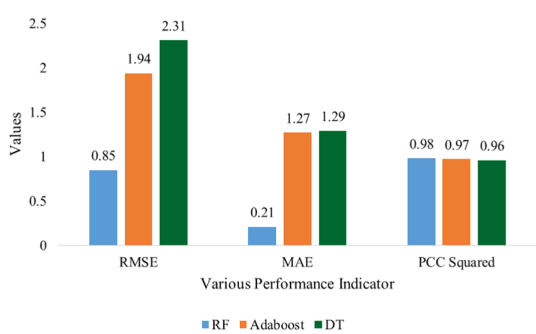

Figure 4. Predictive results of the training datasets.

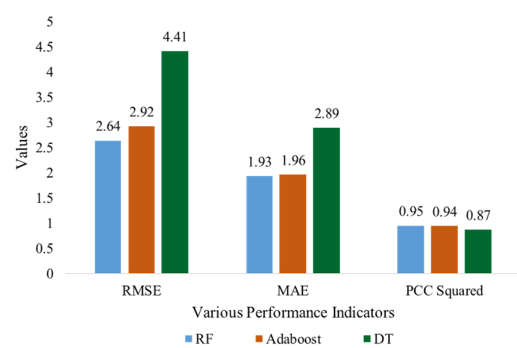

Figure 5. Predictive results of the testing datasets.

8. Monte Carlo simulation method

8.1. General overview

The Monte Carlo (MC) simulation is applied as a probability distribution simulation in various real-world problems [53-54]. This technique is employed as a computerized computation in order to identify different issues such as the uncertainty quantification and risk prediction in various models, such as assessment, prediction, and evaluation in various engineering areas including the finance, project management, decision making, etc. In order to obtain the probabilistic estimation, the MC simulation is manipulated on the repetitive random sampling [55-56]. Each model that predicts the outcomes requires an investigative set of hypotheses associated with the real-world issues and evaluation of the expected values build on the data [55].

Table 2. Spearman's correlation coefficient for the input datasets.

	UCS (MPa)	BTS (MPa)	Brittleness value (S_{20})	Sievers' J value (Sj)
UCS (MPa)	1			
BTS (MPa)	0.575853	1		
Brittleness value (S_{20})	-0.13256	-0.38822	1	
Sievers' J value (S_j)	-0.53985	-0.57547	-0.04251	1

The MC simulation has two important purposes: first, is the quantitative testing of variation and uncertainty, and the second one is the parameter investigation influencing the uncertainty, variation, and their proportion. In contrast to the traditional method, the MC simulation uses an extent of the estimated indicators as an input indicator, and then the system returns an extent of the output indicator for a comprehensive data. Hence, the MC simulation generates a more pragmatic sketch of the probabilistic distribution prediction models. In this technique, a random value is taken for every input indicator contingent on the range of the output indicator, and subsequently, the output datasets are estimated based on these random values. At last, the outcome of the MC simulation is noted. This process is continuously repeated using different random numbers. The mechanism is reproduced by 10000 times. Later on, in the assessment and evaluation of the simulation process, several consequences are generated as an output, which can be used to express the product [57-61].

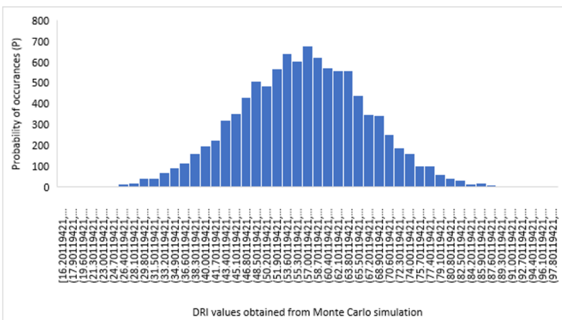

Figure 6. Statistical summary of the DRI histogram obtained from the MC simulation.

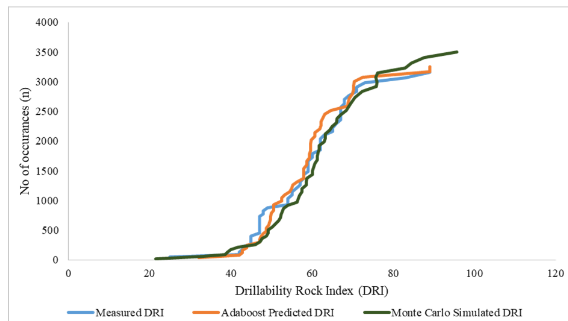

Figure 7. Comparison of the measured, Adaboost predicted, and MC simulated DRI.

8.2. MC Modelling

In this work, an approach of MC was established in order to find a good relationship between the best RF model and the MC simulation. The simple random sampling and Latin hypercube are the two main types of the MC simulation. In order to assure that all the possible combinations are arranged in a stochastic manner, 10000 repetitions were considered during the work meaning that 10000*3-dimension simulations between the input parameters were employed during the MC approach. The correlation between different input parameters could be taken from Table 2. Different simulations are made in order to obtain a statistical representation in this work. The likelihood of occurrence is a criterion function set in the evaluation of the probabilistic model. Figure 6 shows the statistical summary of DRI histogram to the probability of occurrences obtained from the 10000 iteration MC simulation. Figure 7 shows that the measured values of DRI are in the range of 25–89, and the RF predicted values are in the range of 32–89, whereas the MC simulation gives a DRI output of 21.42–93.64, indicating the acceptable measure level of the DRI simulation. A simulation probability of 0.000005 was tied up to recognize the range of DRI. Therefore, all the possible various situations were investigated by the MC simulation. This stochastic approach can be considered as a corner stone for the risk management and probabilistic investigative design of DRI.

9. Conclusions

From the above discussion, the ensemble methods have a great potential for DRI prediction. This work included the performance of three ensemble methods including DT, Adaboost, and RF. Moreover, three different performance indicators were selected in this work in order to check the accuracy of each model. The execution of the ensemble learning method improves its reliability. The proposed ensemble models do not have any well-known specific limitations. However, these models are required to be tested on the field with the real-time drilling data in order to achieve more insight into their performance in the practical scenarios. With the speedy advancement in the sensor-based knowledge acquisition, new measurements could also be available in the future mining, civil, geological, petroleum, and geotechnical engineering projects. This will require re-training of the proposed models with newer training datasets. More research work is required in order to understand DRI with the real-time economic evaluation scenarios.

In short, the following consequences/outcome can be taken from this research work:

i) The constructed simple linear equation showed its ability to measure the value of DRI with 98% accuracy in the training datasets and 95% accuracy in the testing datasets.

ii) The RF model shows a good predictive result for both the training and testing databases among the other ensemble learning models. The RF model has the lowest error rate than the AdaBoost and DT methods.

iii) The range of DRI values simulated by the MC simulation was 25–89 and the RF predicted values were in the range of 32–89, whereas the MC simulation gave a DRI output of 21.42–93.64, demonstrating that the proposed MC simulation could simulate the value of DRI very well by different types of rock databases.

iv) It is highly recommended that the drilling field should be properly investigated before making any decision. Also, it is noteworthy that the equations and models proposed during the work to predict DRI are only related to the current rock engineering issue, and cannot be applied directly to any other rock engineering problem. However, in order to tackle the alternative rock designing and planning projects, the presented developed techniques should be considered a foundation and should be re-evaluated, re-analyzed, and even re-addressed.

Acknowledgments

The author wishes to express their deep gratitude to Dr. Niaz Muhamamd Shahani, China University of Mining and Technology for fruitful discussions and for reading the early version of the draft.

Conflict of interest

The author has no conflict of interest.

Funding

This was an academic research work and no funding source was available.

Appendix 1. Drillability database

S. NO	UCS (MPa)	BTS (MPa)	Brittleness value (S_{20})	Sievers' J value (S_j)	DRI
1	99.19	7.42	43.86	85.7	54
2	177.8	7.66	50.32	7.2	48
3	206.4	13.82	30.02	3.8	25
4	112	10.5	55.87	10.8	57
5	131.5	11.2	64.93	10.7	67
6	156	9.5	61.97	9	62
7	121	11	70.13	14.2	73
8	182	13.5	52.16	3.6	47
9	118	8	69.23	5.6	68
10	52.74	8.69	44.91	82.6	55
11	57.66	6.57	54.38	85.2	65
12	23.43	3.61	82.83	52.3	89
13	45.67	5.4	58.44	52.5	67
14	110.39	10.06	46.85	41	54
15	134.72	5.87	50.43	50.2	59
16	149.87	6.22	53.64	55.4	62
17	169.89	6.75	51.52	48.1	60
18	82	9.1	37.2	41	45
19	83.2	11.8	37.86	41	45
20	75.5	9.3	35.36	30.8	42
21	64.61	5.43	50.76	59.8	59
22	59.25	4.26	60.05	80.1	71
23	61.6	7.2	50.86	52.1	58
24	69.2	6.95	41.25	40.5	47
25	77.8	8.6	37.52	35.1	43
26	73.4	7.7	40.11	42.5	47
27	59.25	4.26	60	80	71

Continuous of Appendix 1

28	61.6	7.2	50.86	52.1	58
29	64.61	5.43	50.76	59.8	59
30	69.2	6.95	41.25	40.5	47
31	77.8	8.6	37.52	35.1	43
32	73.4	7.7	40.11	42.5	47
33	91.55	5.5	36.7	81	47
34	116	9.1	60	53	69
35	128.91	8.67	55.32	87	66
36	151.07	9.26	52.76	63	63
37	88.33	7.08	48.2	51	56
38	23.43	3.61	82.83	52.3	89
39	45.67	5.4	58.44	52.5	67
40	52.74	8.69	44.91	82.6	55
41	57.66	6.57	54.38	85.2	65
42	82	9.1	37.2	41	45
43	83.2	11.8	37.86	41	45
44	75.5	9.3	35.36	30.8	42
45	110.39	10.06	46.85	41	54
46	134.72	5.72	50.43	50.2	59
47	149.87	6.22	53.64	55.4	62
48	169.89	6.75	51.52	48.1	60
49	51	3.28	77.15	41.23	83
50	55.32	5.96	48.19	16.11	49
51	118	8	69.23	5.6	68
52	131.5	11.2	64.93	10.7	67
53	156	9.5	61.97	9	62
54	177.8	7.66	50.32	7.2	48
55	112	10.5	55.87	10.8	57
56	182	13.5	52.16	3.6	47
57	206.4	13.82	30.02	3.8	25

References

[1]. Selmer-Olsen, R. and Lien, R. (1960). Bergartens borbarhet og sprengbarhet. Teknisk Ukeblad, 34, 3-11.

[2]. Hossain, M.E. and Al-Majed, A.A. (2015). Fundamentals of sustainable drilling engineering. John Wiley & Sons.

[3]. Eren, T. and Ozbayoglu, M.E. (2010). Real-time optimization of drilling parameters during drilling operations. In SPE oil and gas India conference and exhibition. Society of Petroleum Engineers.

[4]. Payette, G.S., Spivey, B.J., Wang, L., Bailey, J.R., Sanderson, D., Kong, R. and Eddy, A. (2017). A Real-Time Well-Site based Surveillance and Optimization Platform for Drilling: Technology, Basic Workflows and Field Results. In SPE/IADC Drilling Conference and Exhibition. Society of Petroleum Engineers.

[5]. Osgouei, R.E. (2007). Rate of Penetration Estimation Model for Directional and Horizontal Wells. Master's Thesis, Middle East Technical University, Ankara, Turkey.

[6]. Hoseinie, S.H., Ataei, M. and Mikaeil, R. (2019). Effects of microfabric on drillability of rocks. Bulletin of Engineering Geology and the Environment. 78 (3): 1443-1449.

[7]. Soleimani, M. (2017). Well performance optimization for gas lift operation in a heterogeneous reservoir by fine zonation and different well type integration. Journal of Natural Gas Science and Engineering. 40: 277-287.

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

[9]. Shad, H.I.A., Sereshki, F., Ataei, M. and 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.

[10]. Hoseinie, S.H., Ataei, M. and Mikaiel, R. (2012). Comparison of some rock hardness scales applied in drillability studies. Arabian Journal for Science and Engineering. 37 (5): 1451-1458.

[11]. Kahraman, S., Balcı, C., Yazıcı, S. and Bilgin, N. (2000). Prediction of the penetration rate of rotary blast hole drills using a new drillability index. International Journal of Rock Mechanics and Mining Sciences. 37 (5): 729-743.

[12]. Altindag, R. (2002). The evaluation of rock brittleness concept on rotary blast hold drills. Journal of the Southern African Institute of Mining and Metallurgy. 102 (1): 61-66.

[13]. Bilgin, N. and Kahraman, S. (2003, June). Drillability prediction in rotary blast hole drilling. In Proc. 18th Int. Mining Congress and Exhibition of Turkey, Antalya, Turkey (pp. 177-182).

[14]. Kahraman, S., Bilgin, N. and 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.

[15]. Grima, M.A. and Babuška, R. (1999). Fuzzy model for the prediction of unconfined compressive strength of rock samples. International Journal of Rock Mechanics and Mining Sciences. 36 (3): 339-349.

[16]. Hoseinie, S.H., Ataei, M. and Osanloo, M. (2009). A new classification system for evaluating rock penetrability. International Journal of Rock Mechanics and Mining Sciences. 46 (8): 1329-1340.

[17]. Moradi, H., Bahari, M.H., Sistani, M.B.N. and Bahari, A. (2010). Drilling rate prediction using an innovative soft computing approach. Scientific Research and Essays. 5 (13): 1583-1588.

[18]. Nandi, A.K. and Davim, J.P. (2009). A study of drilling performances with minimum quantity of lubricant using fuzzy logic rules. Mechatronics. 19 (2): 218-232.

[19]. Hashmi, K., Graham, I.D. and Mills, B. (2000). Fuzzy logic-based data selection for the drilling process. Journal of Materials Processing Technology. 108 (1): 55-61.

[20]. Khandelwal, M. and Armaghani, D.J. (2016). Prediction of drillability of rocks with strength properties using a hybrid GA-ANN technique. Geotechnical and Geological Engineering. 34 (2): 605-620.

[21]. Feng, X. (1995). A neural network approach to comprehensive classification of rock stability, blastability and drillability. International Journal of Surface Mining and Reclamation. 9 (2): 57-62.

[22]. Gamal, H., Elkatatny, S. and Abdulraheem, A. (2020, November). Rock Drillability Intelligent Prediction for a Complex Lithology Using Artificial Neural Network. In Abu Dhabi International Petroleum Exhibition & Conference. Society of Petroleum Engineers.

[23]. Fattahi, H. and Bazdar, H. (2017). Applying improved artificial neural network models to evaluate drilling rate index. Tunneling and Underground Space Technology. 70: 114-124.

[24]. Wang, L.N. and Feng, X.T. (1993, January). Comprehensive Classification Iof Rock Stability, Blastability and Drillability based on Neural Networks. In The 34th US Symposium on Rock Mechanics (USRMS). American Rock Mechanics Association.

[25]. Gan, C., Cao, W., Wu, M., Chen, X., Lu, C., Hu, Y. and Wen, G. (2016). Intelligent Nadaboost-ELM modeling method for formation drillability using well logging data. Journal of Advanced Computational Intelligence and Intelligent Informatics. 20 (7): 1103-1111.

[26]. Gan, C., Cao, W., Wu, M., Chen, X., Hu, Y., Wen, G. and Ding, H. (2017). An online modeling method for formation drillability based on OS-Nadaboost-ELM algorithm in deep drilling process. IFAC-PapersOnLine. 50 (1): 12886-12891.

[27]. Tewari, S., Dwivedi, U.D. and Biswas, S. (2021). A Novel Application of Ensemble Methods with Data Resampling Techniques for Drill Bit Selection in the Oil and Gas Industry. Energies. 14 (2): 432.

[28]. Li, S., Zhang, J., Wu, S., Chen, W., Chen, D., Li, X. and Wang, H. (2019, September). Prediction of Rate of Penetration based on Random Forest in Deep Well. In International Conference on Inforatmion technology in Geo-Engineering. (pp. 517-526). Springer, Cham.

[29]. Mikaeil, R., Haghshenas, S.S. and Hoseinie, S.H. (2018). Rock penetrability classification using artificial bee colony (ABC) algorithm and self-organizing map. Geotechnical and Geological Engineering. 36 (2): 1309-1318.

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

[31]. Saeidi, O., Torabi, S.R., Ataei, M. and Rostami, J. (2014). A stochastic penetration rate model for rotary drilling in surface mines. International Journal of Rock Mechanics and Mining Sciences. 68: 55-65.

[32]. Sagi, O. and Rokach, L. (2018). Ensemble learning: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 8 (4): e1249.

[33]. Kotsiantis, S.B. (2013). Decision trees: a recent overview. Artificial Intelligence Review, 39(4), 261-283.

[34]. Breiman, L. (2001). Random forests. Machine learning. 45 (1): 5-32.

[35]. Freund, Y. and Schapire, R.E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. Journal of computer and system sciences. 55 (1): 119-139.

[36]. Lyubartsev, A.P., Martsinovski, A.A., Shevkunov, S.V. and VorontsovVelyaminov, P.N. (1992). New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles. The Journal of chemical physics. 96 (3): 1776-1783.

[37]. Iba, Y. (2001). Extended ensemble Monte Carlo. International Journal of Modern Physics C, 12(05), 623-656.

[38]. Yenice, H. (2019). Determination of Drilling Rate Index Based on Rock Strength Using Regression Analysis. Anais da Academia Brasileira de Ciências, 91 (3).

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

[40]. Azizi, A., Shafaei, S.Z., Rooki, R., Hasanzadeh, A. and Paymard, M. (2012). Estimating of gold recovery by using back propagation neural network and multiple linear regression methods in cyanide leaching process.

[41]. Williamson, D.F., Parker, R.A. and Kendrick, J.S. (1989). The box plot: a simple visual method to interpret data. Annals of internal medicine. 110 (11): 916-921.

[42]. Witten, I.H. and Frank, E. (2005). Data Mining — Practical Machine Learning Tools and Techniques, Second Edition. Elsevier, Amsterdam.

[43]. Friedl, M.A. and Brodley, C.E. (1997). Decision tree classification of land cover from remotely sensed data. Remote sensing of environment. 61 (3): 399-409.

[44]. Pal, M. and Mather, P.M. (2003). An assessment of the effectiveness of decision tree methods for land cover classification. Remote sensing of environment. 86 (4): 554-565.

[45]. Zhang, J., Li, D. and Wang, Y. (2020). Toward intelligent construction: prediction of mechanical properties of manufactured-sand concrete using tree-based models. Journal of Cleaner Production. 258: 120665.

[46]. Sun, W. and Gao, Q. (2019). Exploration of energy saving potential in China power industry based on Adaboost back propagation neural network. Journal of Cleaner Production. 217: 257-266.

[47]. Hong, H., Liu, J., Bui, D.T., Pradhan, B., Acharya, T.D., Pham, B.T. and Ahmad, B.B. (2018). Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). Catena. 163: 399-413.

[48]. Maniruzzaman, M., Rahman, M.J., Al-MehediHasan, M., Suri, H.S., Abedin, M.M., El-Baz, A. and Suri, J.S. (2018). Accurate diabetes risk stratification using machine learning: role of missing value and outliers. Journal of medical systems. 42 (5): 1-17.

[49]. Zhang, J., Xu, J., Hu, X., Chen, Q., Tu, L., Huang, J. and Cui, J. (2017). Diagnostic method of diabetes based on support vector machine and tongue images. BioMed research international, 2017.

[50]. Yang, P., Hwa Yang, Y., B Zhou, B. and Y Zomaya, A. (2010). A review of ensemble methods in bioinformatics. Current Bioinformatics. 5 (4): 296-308.

[51]. Breiman, L. (1996). Bagging predictors. Machine learning. 24 (2): 123-140.

[52]. Liang, W., Sari, A., Zhao, G., McKinnon, S.D. and Wu, H. (2020). Short-term rockburst risk prediction using ensemble learning methods. Natural Hazards, 104 (2): 1923-1946

[53]. Wattimena, R.K., Kramadibrata, S., Sidi, I.D., Arif, I. and Azizi, M.A. (2012, January). Probabilistic analysis of single bench using new slope stability curves. In ISRM Regional Symposium-7th Asian Rock Mechanics Symposium. International Society for Rock Mechanics and Rock Engineering.

[54]. Azizi, M.A., Kramadibrata, S., Wattimena, R.K. and Sidi, I.D. (2013). Probabilistic analysis of physical models slope failure. Procedia Earth and Planetary Science, 6: 411-418.

[55]. Solver, F. (2010). Premium solver platform. User Guide, Frontline Systems.

[56]. EPA, U. (1997). Environmental Protection Agency. Guiding principles for Monte Carlo analysis. EPA/630/R-97/001.

[57]. Sari, M. and Ataei, M. (2012). Development of an empirical model for predicting the effects of controllable blasting parameters on fly-rock distance in surface mines. International Journal of Rock Mechanics and Mining Sciences, 52, 163-170.

[58]. Steinfeld, B., Scott, J., Vilander, G., Marx, L., Quirk, M., Lindberg, J. and Koerner, K. (2015). The role of lean process improvement in implementation of evidence-based practices in behavioral health care. The Journal of Behavioral Health Services and Research, 42 (4): 504-518.

[59]. Mahdiyar, A., Hasanipanah, M., Armaghani DJ et al A Monte Carlo technique in safety assessment of slope under seismic condition. Eng Comput. 2017. https ://doi.org/10.1007/s0036 6-016-0499-1

[60]. Bianchini, F. and Hewage, K. (2012). Probabilistic social cost-benefit analysis for green roofs: A lifecycle approach. Building and Environment, 58: 152-162.

[61]. Dunn, W.L. And Shultis, J.K. Monte Carlo methods for design and analysis of radiation detectors. Radiat Phys Chem. 2009; 78:852–858. https ://doi.org/10.1016/j.radph ysche m.2009.04.030.

رویکرد احتمالی برای پیش بینی شاخص نرخ حفاری با استفاده از روش یادگیری گروه

محمد کامران*

گروه مهندسی معدن، موسسه فناوری باندونگ، کوتا باندونگ، اندونزی

ارسال ۲۰۲۱/۰۴/۰۷، پذیرش ۲۰۲۱/۰۵/۰۴

* نویسنده مسئول مکاتبات: muhdkamran62@gmail.com

چکیده:

قابلیت حفاری یکی از موضوعات مهم مهندسی سنگ است. شاخص سرعت حفاری (DRI) ابزاری مهم در تجزیه و تحلیل قابلیت حفاری سنگ‌ها است. محققان تلاش‌های زیادی برای همبستگی و ارزیابی DRI سنگها انجام داده‌اند. روش‌های یادگیری گروهی شامل درخت تصمیم (DT)، تقویت انطباقی (AdaBoost) و جنگل تصادفی (RF) در این تحقیق به منظور پیش بینی DRI سنگ‌ها به کار رفته است. یک پایگاه داده با قابلیت تمرین با چهار پارامتر در این کار وارد شده است. رابطه بین پارامترهای ورودی و DRI با استفاده از تحلیل رگرسیون ساده برقرار شده است. به منظور آموزش مدل، خصوصیات مکانیکی مختلف سنگها شامل مقاومت فشاری تک محوری (UCS)، مقاومت کششی برزیلی (BTS)، آزمون شکنندگی (S20) و اندازه سرمته (Sj) به عنوان ورودی در نظر گرفته شده است. پایگاه داده اصلی DRI به طور تصادفی به مجموعه آموزش و آزمون با روش نمونه گیری به نسبت ۸۰ به ۲۰ تقسیم می‌شود. الگوریتم‌های مختلفی توسعه داده شده است، و در نتیجه، چندین روش برای پیش بینی DRI نمونه‌های سنگ ارائه شد. عملکرد مدل نشان داده است که روش RF، DRI را با دقت بالایی پیش بینی می‌کند.علاوه بر این، شبیه سازی‌های مونت کارلو نشان می‌دهد که این روش در پیش بینی توزیع احتمال DRI قابل اطمینان‌تر است. بنابراین، مدل پیشنهادی را می‌توان برای مدیریت ریسک پایداری و طراحی تحقیقاتی DRI بکار گرفت.

کلمات کلیدی: شاخص سرعت حفاری، یادگیری گروهی، پیش‌بینی، قابلیت حفاری، احتمال.