Shahrood University of Technology

Journal of Mining and Environment (JME)

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

Iranian Society of Mining Engineering (IRSME)

Grade Prediction of an Indian Iron Ore Mining using Machine Learning and Ensemble Models

Tapan Dey[1*], and Gopinath Samanta[2]

1. Research Scholar, Mining Engineering, NIT Rourkela, Kolkata, India
2. Senior Technologist, Tata Steel Ltd., Jamshedpur, India

Article Info

Received 28 June 2025

Received in Revised form 14 September 2025

Accepted 2 November 2025

Published online 2 November 2025

DOI: 10.22044/jme.2025.16430.3208

Keywords

ML

Kriging

XGBoost

Random Forests

Estimation

Abstract

Accurate grade prediction is an important step in the mining planning process. Various methods, namely the Inverse Distance Method and Kriging, are widely used. The application of machine learning is a new development in the grade estimation technique. The present study focused on the application of XGBoost, Random Forests (RFs), Multi Layer Perceptron (MLP), and Gradient Boosting Regression (GBR) models to predict iron ore grades in an Indian mine. An ensemble model was also applied to obtain a more stable grade prediction in the deposit. Models were trained using 4,112 sample data, which have spatial coordinates (east, north, and altitude) and iron grades. The dataset was divided into two parts: 80% (3,289 samples) of the data was used for model training, and 20% (823 samples) was used for model testing. The performance of the models was assessed through the coefficient of determination (R^2), Mean Squared Error (MSE), and Mean Absolute Percentage Error (MAPE). The results show that the XGBoost model performs better in the estimation process when compared with other methods, such as RFs, GBR, and MLP. The XGBoost model produced R^2 of 0.77, MSE of 2.87, and MAPE of 1.8%. The findings indicate that the XGBoost model is effective for predicting iron ore grades in this type of deposit. However, considering geological uncertainty, the application of an ensemble model may be beneficial for grade prediction in an iron deposit.

1. Introduction

Mine production planning, scheduling, and financial analysis heavily rely on the block grade model [1-3]. Exploratory drilling is conducted in limited locations within the ore body where reserve estimation is to be performed. Therefore, a robust estimation technique is crucial for determining the potential reserve of the deposit. Accurate estimation is challenging as the formation of ore deposition is a complex process that is not always fully understood [4]. Most mine planning software utilizes conventional grade estimation techniques. For example, kriging is an established geostatistical method [5]. A geostatistical model operates on the deposit structure by calculating the variogram. However, due to limited sample points and the complexity of the deposit, structural analysis can be a difficult task. These methods require expertise, considerable time, and certain assumptions [6-7]. The Inverse Distance Weighting (IDW) method assigns weights to samples based on their distance, giving more influence to those closer to the point of interest [8].

With the rapid advancement of technology, the use of existing methods is continuously

Corresponding author: tapan.besu@gmail.com (T. Dey)

questioned. Concurrently, mining method selection increasingly emphasizes the limitations and prediction accuracy of the estimation process. Nowadays, technological progress has yielded impressive results across diverse fields, ranging from mining to space exploration. The increasing power of computing systems has fueled the significant growth of soft computational techniques. Today, machine learning and artificial intelligence are rapidly becoming dominant forces across all fields [9-11]. Consequently, research in a range of industries, including mining, is exploring the application of artificial intelligence (AI) and machine learning (ML) for solving engineering problems with nonlinear relationships [12-13]. Various machine learning models, popular among researchers, include random forests (RFs), multilayer perceptron (MLP), XGBoost, gradient boosting regression (GBR), and many others [14]. The abundance of opportunities has attracted many researchers to employ data-driven approaches for precise prediction without any pre-assumptions.

Previous research on the Darreh-Ziarat iron ore deposit incorporates auxiliary variables, such as magnetometry-derived susceptibility, into co-kriging and co-simulation. This significantly reduces estimation error and improves mineral resource classification [15]. A recent study on the Madan Bozorg copper mine demonstrates that 3D geostatistical modeling of induced polarization and resistivity data, integrated with drilling, can reliably delineate mineralization zones and optimize drilling locations [16]. In the Dalli gold deposit, machine learning approaches such as SVM, BPNN (Back-Propagation Neural Network), and multivariate regression have been applied for grade prediction. In this study, SVM demonstrates the highest accuracy and reliability [17]. A case study at the Abassabad copper mine showed that integrating drilling with induced polarization and resistivity data through geostatistical methods improves resource estimation and allows optimization of borehole locations [18]. Geostatistical modeling of seismic velocity data in a carbonate oilfield of SW Iran has shown that integrating Vp–Vs cubes with acoustic impedance inversion improves pore pressure and fracture pressure estimation accuracy [19]. A three-dimensional stope layout optimization framework is developed that explicitly incorporates grade variability through geostatistical simulation, thereby enabling more robust and risk-informed design decisions in comparison to conventional kriging-based approaches [20]. Sequential Gaussian Simulation (SGS) better captures ore grade uncertainty than Ordinary Kriging (OK), resulting in more reliable mine plans and a higher net present value. Sequential Gaussian conditional simulation outperforms kriging in grade estimation for the Sungun copper deposit. It increases mineable ore and improves net present value by 40% [21, 22]. Gaussian copulas provide more accurate and precise copper grade estimates than ordinary and indicator kriging in the Sungun deposit by accounting for both distance and sample values [23].

A probabilistic model for grade reconciliation in open pit mining demonstrates that systematic uncertainty and random variability are the primary causes of discrepancies between estimated and actual grades [24]. A correction-factor model is proposed by addressing systematic, inherent, and statistical uncertainties in grade estimation, which improved reconciliation from 10% to 50% in an Iranian iron ore open-pit mine [25]. A hybrid extreme learning machine–particle swarm optimization method effectively improves iron ore grade estimation using borehole data from the Choghart deposit in Iran [26]. AI techniques using Python to optimize borehole drilling and iron grade estimation in the Choghart deposit achieved over 91% accuracy with low prediction errors [27].

In this study, iron ore grade estimation using XGBoost was compared with traditional kriging on drillhole composites. The study found that XGBoost produced a wider grade range and partially reduced the smoothing effect relative to kriging, although both methods still yielded smoother results than the original samples. This approach is effective, but it remains a single-model approach. In the Singhbhum-Keonjhar-Bonai iron ore belt, Fe grade variability in BIF-hosted deposits was modeled using ordinary kriging (OK) and sequential Gaussian simulation (SGS). The study confirmed lognormal grade distributions with spherical variogram structures, where OK provided conditionally unbiased estimates and SGS generated multiple realizations that reproduced the variability of the original data [28].

While most literature describes machine learning models with conventional geostatistical techniques, the present study introduces a rigorous spatial framework and leverages ensemble models. It also delivers a direct comparison with the conventional reserve estimation technique, kriging. Considering the limited research on Indian iron ore deposits, this study will add significant value to the Indian iron mining context.

In this study, various ML models are developed for predicting the iron ore grade of a mine. The prediction model utilizes X, Y, and Z coordinates for grade estimation. The study primarily focuses on the estimation of iron ore (Fe) grade utilizing various ML models, as well as ensemble models. Python programming is used for the development of ML models.

2. Materials and Methods

2.1. Geological Parameters of Deposit

The borehole data was collected from the eastern portion of the Bonai-Keonjhar (BK) deposit in the Indian state of Odisha. The main rock formations in the region consist of metamorphosed sandy and clay-rich sediments, with occasional lens-shaped calcareous bands. The iron ore-bearing and associated rocks in the area belong to a Precambrian sedimentary sequence that hosts diverse iron ore types, including banded iron formation (BIF), banded hematite jasper (BHJ), detrital iron deposits (DID), and brecciated iron deposits (BID) [29]. The dominant iron oxides are hematite, magnetite, martite, and specularite, while silica occurs mainly as quartz, jasper, and chert. Structurally, the strata display a synclinorial pattern trending NNE–SSE with a gentle plunge towards the NNE. The mineralization is largely confined to BHJ on the northern, western, and southern margins, whereas the eastern part is mostly overlain by laterite.

The effectiveness of machine learning models in predicting iron ore grades is strongly influenced by the geological characteristics of the deposit. Deposits that are stratified and uniform in mineralogy, such as those hosted in Banded Iron Formation (BIF), generally provide better predictive outcomes owing to their spatial continuity and stable grade distribution. In contrast, deposits with structural complexity, diverse mineralogy, irregular gangue content, or skewed grade patterns create higher levels of uncertainty in model performance. Additional factors, including sampling density, weathering variations with depth, and overall data quality, also affect reliability, highlighting the importance of advanced modeling techniques and geologically informed feature selection.

2.2. Data Collection and General Parameters

This study analyzes iron ore grade data obtained from seventy-one (71) exploratory drill holes. The depth of the holes varies from 7 m to 125 meters, and the average depth is 63 meters. The exploratory borehole data is given in Figure 1, and the plan view of the drill hole data in the X-Y plane (for z = 798.5) is shown in Figure 2.

2.3. Data Processing

2.3.1. Outlier detection and treatment

Outliers were identified using statistical methods such as boxplots and interquartile range (IQR) analysis as shown in Figure 3. The extreme values that did not represent geological or sampling reality were capped using the upper quartile (95%) and lower quartile (5%) to minimize their influence on model training.

2.3.2. Handling of missing data

The datasets were examined to identify missing data or incomplete records. There were only a few missing data points found, and these data points were removed.

2.3.3. Final data quality check

After preprocessing, statistical summaries, correlation matrices, and visualizations were used to confirm data integrity. The cleaned and transformed datasets were split into training and testing subsets for performing a machine learning model.

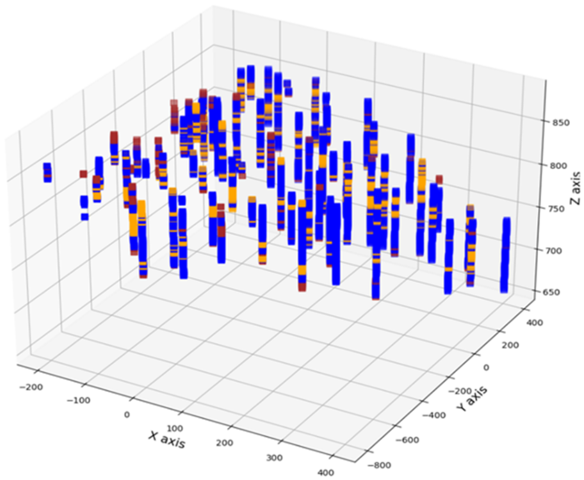


Figure 1. Drill holes data in three dimensions

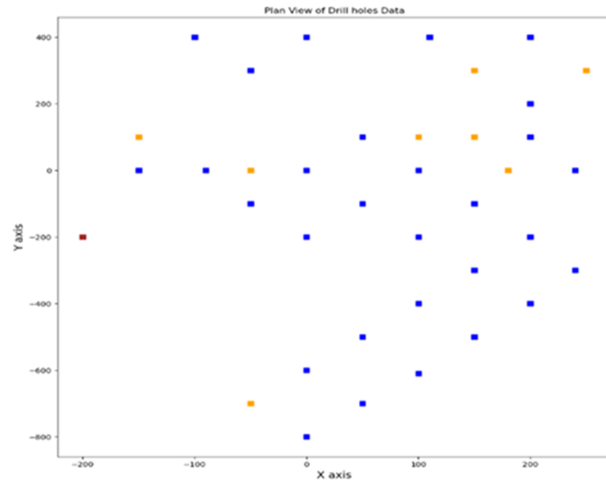
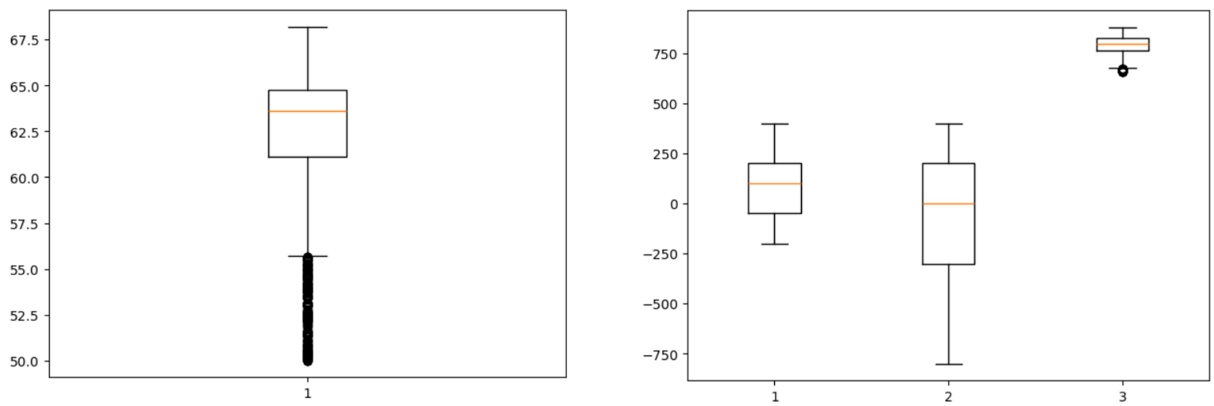


Figure 2. Plan view of drill holes (at z = 798.5)



(a) **(b)**

Figure 3. Box plot of (a) iron grade, and (b) coordinates

2.4. Machine Learning Models

Machine learning models are employed for predicting ore grades due to several key advantages they hold over conventional methods. These include their ability to manage complex, non-linear relationships, improved prediction accuracy, the capability to handle large and heterogeneous datasets, adaptation to new data, and ultimately reduced exploration costs. In this study, ML models such as XGBoost, RFs, GBR, and MLP have been utilized for ore grade prediction.

2.4.1. Random forests (RFs)

RFs are a widely used ensemble learning method offering high prediction accuracy and robustness due to their reliance on bagging and random subspace methods [30]. This approach mitigates overfitting by aggregating predictions from multiple samples and considering random features at each split. Random forests effectively handle high dimensionality, provide feature importance measures, and are readily implemented using various libraries [31]. The key steps for regression using RF are as follows:

1. Bootstrap Sampling: It is used to create several datasets (D) by randomly sampling from the original dataset with replacement. For each bootstrap sample of size n, number of bootstrap samples is generated by selecting random samples with replacement.
2. Tree Construction: For each bootstrap sample, a regression tree is generated by recursively splitting the nodes. At each split, a random selection of m features is made from the total p features to identify the best possible split.

3. Aggregation: The final prediction is determined by combining predictions from all number of trees using equation 1.

$$\hat{y} = \frac{1}{B}\sum_{b=1}^{B} T^{(b)}(x) \quad (1)$$

Where are:

$\hat{y}$ − final predicted value,

B − number of bootstrap samples,

$T^{(b)}$ – regression tree.

2.4.2. Extreme Gradient Boosting (XGBoost)

XGBoost efficiently minimizes the loss function iteratively with the help of gradient boosting [32]. It combines trees in a sequence, where trees improve the errors. The loss function and iterative prediction update are given in equations 2 and 3.

$$L = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \quad (2)$$

Where are:

l – loss (e.g., squared error for regression),

$\Omega(f_k)$ – penalizes model complexity.

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + \eta f_t(x_i) \quad (3)$$

Where are:

η – learning rate,

$f_t(x_i)$ – output of the t^{th} tree.

2.4.3. Gradient Boosting Regression (GBR)

GBR is an iterative ensemble method that builds trees in a sequence to minimize the residual of the previous tree using gradient descent. GBR's strength is in modeling non-linear relations and flexibility in optimizing custom loss functions, making it a robust choice for regression tasks. The loss function and negative gradient of loss functions are captured by equations 4 and 5.

$$L = \sum_{i=1}^{n} l(y_i, \hat{y}_i) \quad (4)$$

Where are:

l – loss (e.g., squared error for regression),

y_i – true value,

$\hat{y}_i$ – predicted value.

$$g_i = -\frac{\partial l(y_i, \hat{y}_i)}{\partial \hat{y}_i} \quad (5)$$

Where are:

y_i – true value,

$\hat{y}_i$ – predicted value,

g_i – gradient.

2.4.4. Multi-Layer Perceptron (MLP)

MLP plays a foundational role in machine learning. It is a universal approximator, capable of modeling complex non-linear functions, making them versatile for tasks like classification, regression, and time-series prediction. However, they require careful regularization and hyperparameter tuning to avoid overfitting. MLPs are composed of one input layer, one or more hidden layers, and one output layer [33]. Each layer consists of neurons, and the connections between them are associated with weights. The input x is transformed as it propagates through the network using weights W and biases b, followed by activation functions. Mathematically, the transformation for a layer is given by:

$$z = Wx + b \quad (6)$$

Where are:

z – pre-activation values,

W – network weights,

b – network biases.

$$a = \sigma(z) \quad (7)$$

Where are:

a – activation output,

$\sigma(.)$ – activation function (e.g., ReLU or sigmoid),

z – pre-activation values.

The output layer provides predictions by minimizing a loss function during the training of the model. Training is performed using backpropagation and gradient descent, iteratively updating weights as in equation 8.

$$W_{new} = W - \eta \nabla L \quad (8)$$

Where are:

W – network weights,

η – learning rate.

2.4.5. Ordinary Kriging Method

Geostatistical estimation typically involves three main steps: examining spatial continuity, constructing a variogram model, and carrying out kriging estimation. Spatial continuity describes how sample values are correlated as the distance between them varies. This can be illustrated through scatter or h-scatter plots. The semivariogram serves as a mathematical tool to quantify this continuity within an orebody and is derived using Equation (9).

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (Z(u_i) - Z(u_i + h))^2 \quad (9)$$

Where are:

$\gamma(h)$ – semivariogram value,

$Z(u_i)$ – grade observed at location u_i,

h – represents the separation (lag) distance between two points,

N(h) – indicates the number of data pairs available at that lag

Based on these calculations, an experimental semivariogram is generated. To obtain a smooth and continuous representation, a theoretical model such as spherical, Gaussian, or exponential is fitted to the experimental curve [34]. A representative semivariogram is illustrated in Figure 4. The key parameters derived from the semivariogram are the nugget, sill, and range.

Ordinary Kriging (OK) is a geostatistical interpolation method that relies on semivariogram information to predict values at unsampled locations. By limiting the assumption of stationarity to a local neighbourhood of samples, OK accounts for variations in the local mean [35]. Instead of requiring prior knowledge of the global mean, it enforces an unbiasedness condition, ensuring that the sum of the assigned weights equals 1. A distinguishing characteristic of kriging is that it not only produces an estimated value of the variable but also provides the corresponding estimation variance, thereby quantifying the reliability of the prediction at each location [36]. The estimation process and the expression for minimum variance in OK are given by Equations (10) and (11).

$$Z^*(u_0) = \sum_{i=1}^{n} \lambda_i . Z(u_i) \;\; ; \sum_{i=1}^{n} \lambda_i = 1 \quad (10)$$

Where are:

$Z^*(u_0)$ – estimated grade at the unsampled location u_0,

$Z(u_i)$ – grade at the location u_i,

$[\lambda]$ – weighting factors obtained from the variogram model.

$$\sigma_{u_0}^2 = \sum_{i=1}^{n} \lambda_i * \gamma_{u_i,u_o} + \mu \quad (11)$$

Where are:

$\sigma_{u_0}^2$ – estimation variance at that point at u_0,

$[\lambda]$ – weighting factors obtained from the variogram model,

μ – Lagrange multiplier used in the kriging system,

γ_{u_i,u_o} – covariance between the sample location at u_i and the estimation point u_0.

3. Results and Interpretation

In this study, the dataset was carefully divided, with 20% allocated for testing purposes and the remaining 80% for model training. Random Forest (RFs), XGBoost, Multilayer Perceptron (MLP), and Gradient Boosting Regression (GBR) models were explored considering the less correlation between input and output variables. Kaplan et al. and Atalay et al. [37–38] discussed the spatial correlation between the X, Y, Z coordinates and the Au/Fe grades.

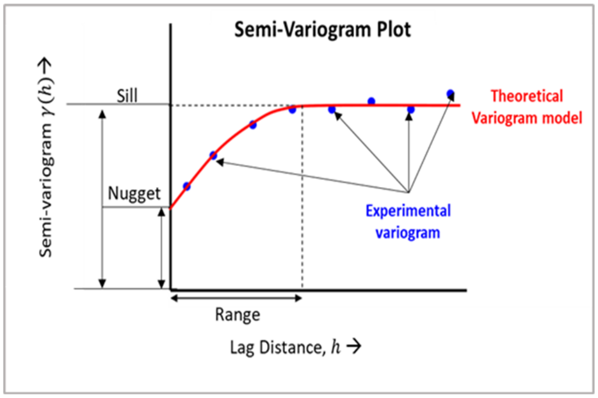


Figure 4. Semi-variogram plot showing experimental and theoretical variogram

3.1. Model Training

In the first step, the training dataset was used to train the models. Error minimization was performed through hyperparameter tuning using the grid search method. The optimal number of estimators and mean squared error (MSE) for RFs, XGBoost, GBR, and MLP trained models were (700, 3.09), (2750, 4.5), (14000, 2.87), and (1500, 4.67), respectively. Hyperparameter tuning plots are shown in Figure 5, and the optimal values of hyperparameters for different ML models are given in Table 1. Computation details for various models are captured in Table 2.

Table 1. Optimal value of the hyper parameters for various ML models

Name of the Model	Name of the Hyper Parameters	Hyper Parameters Value	Description of the Hyper Parameters
Random Forests (RFs)	Estimators	700	Number of trees
	Maximum depth	35	Maximum depth of the tree
	Minimum samples in leaf node	2	Minimum number of data point in each leaf node
Extreme Gradient Boosting (XGBoost)	Estimators	2750	Number of trees
	Learning rate	0.02	Learning rate
	Objective	linear	Linear regression with squared error
Gradient Boosting Regression (GBR)	Estimators	4210	Number of trees
	Learning rate	0.02	Learning rate
Multilayer Perceptron (MLP)	Neurons in hidden layer	400	Number of neurons in the hidden layer
	Activation function	Logistic	Activation function used to get the final output
	Maximum iterations	15000	Maximum number of iterations to train the model

Table 2. Computation details for various Models

Name of the Model	Training time (mins)	Prediction time (mins)	Memory use (%)
Random Forests (RFs)	20	4	11%
Extreme Gradient Boosting (XGBoost)	15	3	9%
Gradient Boosting Regression (GBR)	12	2	8%
Multilayer Perceptron (MLP)	25	5	15%
Kriging	120	10	47%
Ensemble		5	5%

3.2. Model Validation

Once hyperparameter tuning was completed and models were finalized, the iron ore grade of the test data was predicted using the trained models. Then, the predicted grade was compared with the actual value through R^2, MSE, and mean absolute percentage error (MAPE). The XGboost model performed best out of the four models, with an R^2 of 0.77, an MSE of 2.87, and a MAPE of 1.8%. However, other models (RFs, MLP) also performed well. Figures 6 display the grade predictions of the selected models compared to the actual values for the training and testing datasets. It also shows that prediction at a lower grade point (54%) is not stable and provides a wide range of predicted values both for training and testing. The detailed accuracy metrics of all four models are given in Table 3.

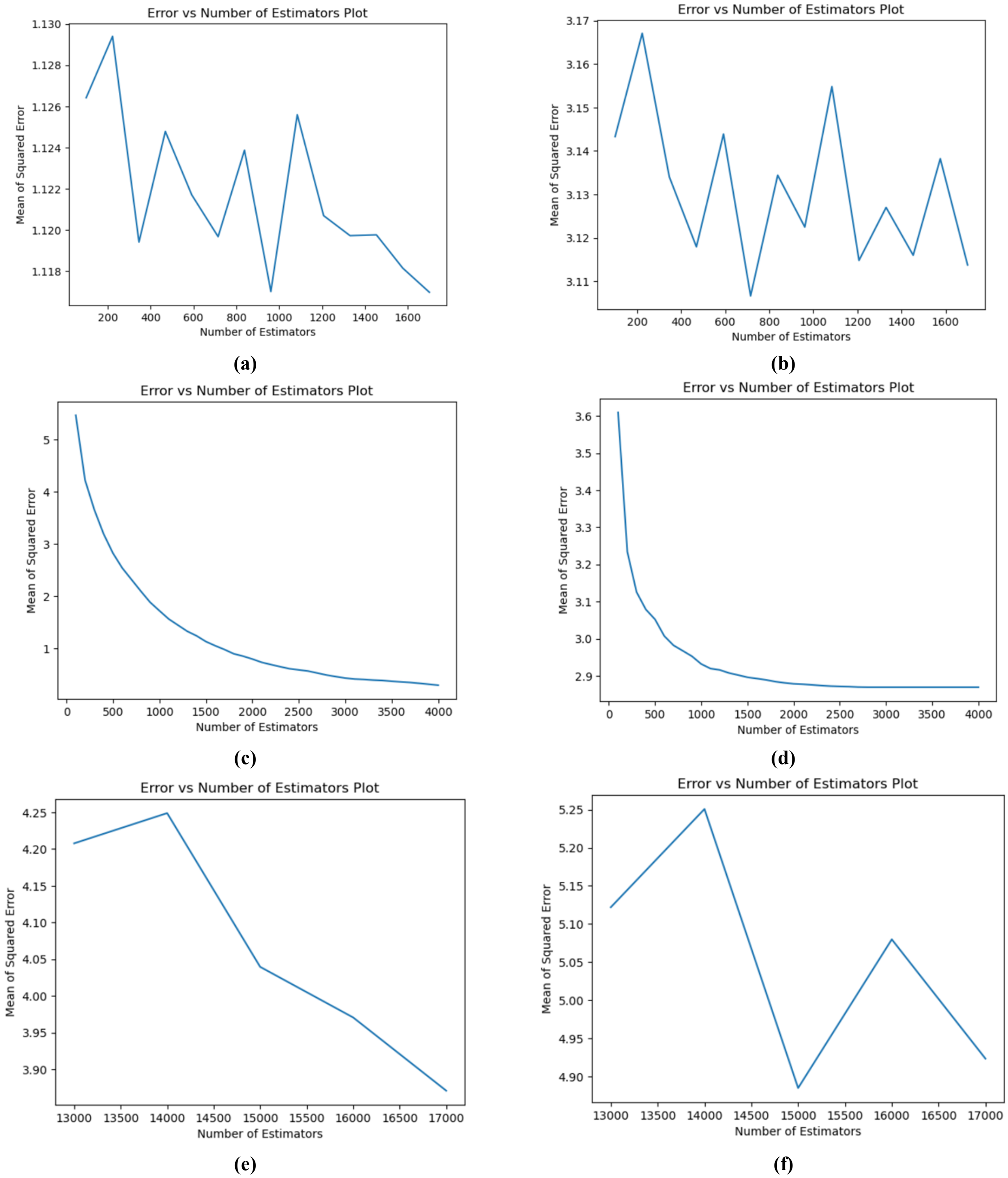
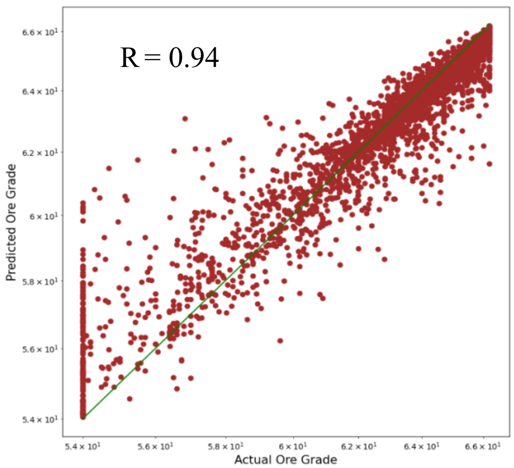
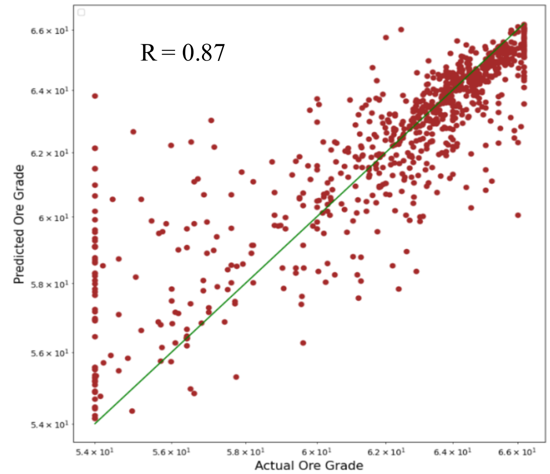


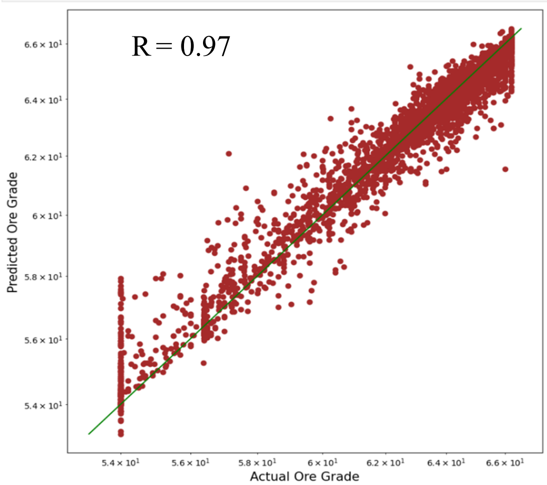
Figure 5. Hyper parameter tuning plots of (a) RFs Training, (b) RFs Testing, (c) XGBoost Training, (d) XGBoost Testing, (e) MLP Training, and (f) MLP Testing



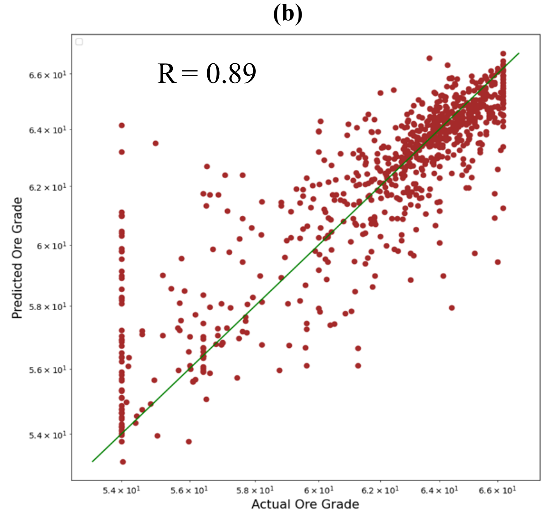
(a)



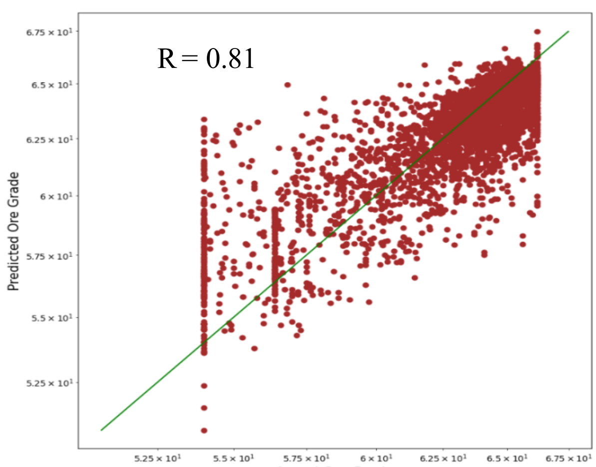
(b)



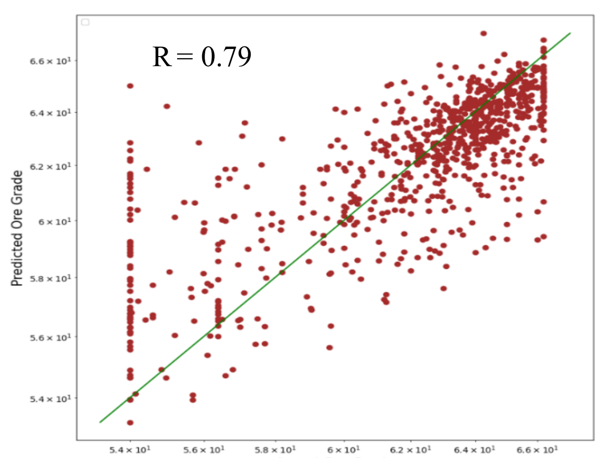
(c)



(d)

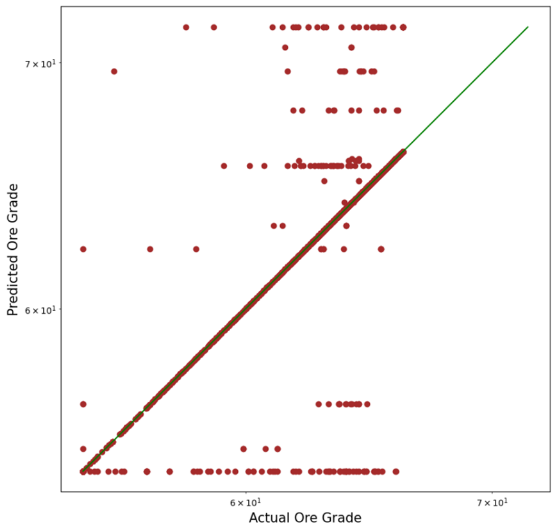


(e)

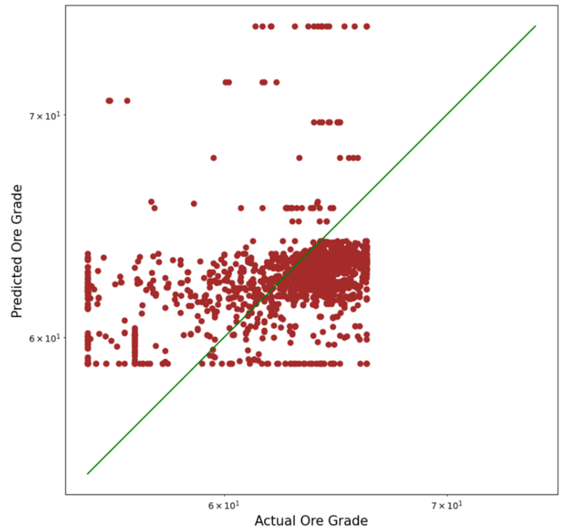


(f)

Figure 6. Actual vs Predicted plot of (a) RFs Training model, (b) RFs Testing model, (c) XGBoost Training model, (d) XGBoost Testing model, (e) MLP Training model, (f) MLP Testing model, (g) Kriging Training model, and (h) Kriging Testing model



(g)



(h)

Continuous of Figure 6. Actual vs Predicted plot of (a) RFs Training model, (b) RFs Testing model, (c) XGBoost Training model, (d) XGBoost Testing model, (e) MLP Training model, (f) MLP Testing model, (g) Kriging Training model, and (h) Kriging Testing model

Table 3. Accuracy metrics for various models

Method	R²	MSE	MAPE
Randon Forest (RFs)	0.75	3.09	1.9%
Extreme Gradient Boosting (XGBoost)	0.77	2.87	1.8%
Gradient Boosting Regression (GBR)	0.57	5.31	2.7%
Multilayer Perceptron (MLP)	0.62	4.67	2.5%
Kriging	0.71	3.24	2.2%
Ensemble Model	0.76	2.95	1.9%

K-fold cross-validation was also conducted to check the bias of the training and testing results. For this exercise, the two best-performing models, XGBoost and RFs, were used. A 5-fold cross-validation exercise was performed. The model performance results are given in Table 4. Since the cross-validation output is close to the results of the train-test model, the ML model is performing efficiently in predicting the ore grade.

Table 4. Accuracy metrics for k-fold cross validation models

Method	Individual MSE scores for each fold	Average MSE across all folds
Randon Forest (RFs)	2.95, 3.39, 3.23, 3.48, 3.41	3.29
Extreme Gradient Boosting (XGBoost)	2.94, 2.96, 3.10, 2.85, 3.05	2.98

3.3. Ensemble models

From individual model performance, it was observed that three models were performing with decent accuracy. Ensemble models are constructed to reduce prediction uncertainty, reduce overfitting, and enhance robustness. They leverage the strengths of different models to create more reliable and generalized predictions [39]. Out of various ensemble models, the weighted average ensemble model was used to determine the final prediction from the individual model predictions. The ensemble model's R^2 was 0.76, MSE was 2.95, and MAPE was 1.9%. Although the R^2 of the ensemble model was reduced by 0.01, the ensemble model may produce more stable and reliable predictions (Figure 7) when compared with the individual model. Comparative predictions from machine learning models and the conventional method kriging are given in Figure 8.

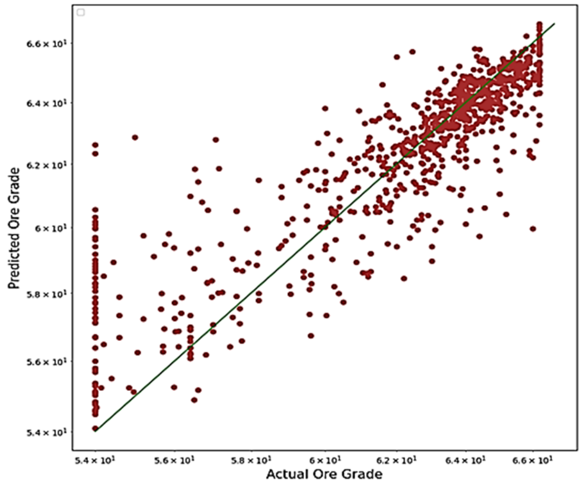


Figure 7. Actual vs Predicted plot for ensemble out for testing data

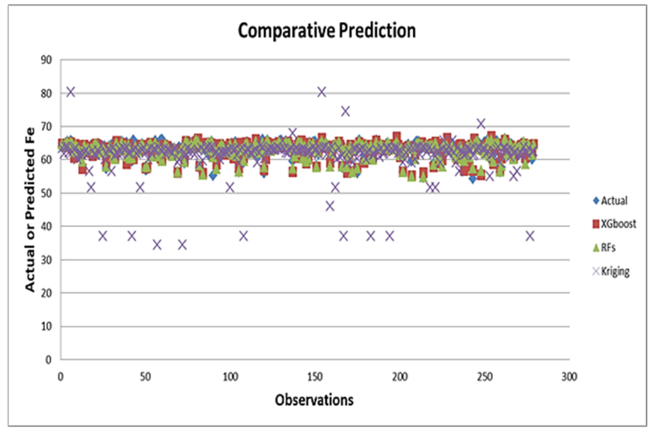


Figure 8. Comparative Predicted value from various models for testing data

4. Conclusions

The present study focuses on the usage of various machine learning (ML) models for estimating the iron ore grade in Indian deposits. Among various ML models, four models—namely, random forests (RFs), XGBoost, gradient boosting regression (GBR), and multi-layer perceptron (MLP)—were trained to predict the iron ore grade in unsampled locations. Rigorous hyperparameter tuning was also performed to achieve the best model accuracy. An ensemble model was developed to obtain a stable prediction with minimum uncertainty. This study shows that RFs, XGBoost, and MLP models produce good accuracy. However, the XGBoost model was found to be superior, having an R^2 of 0.77, a mean squared error (MSE) of 2.87, and a mean absolute percentage error (MAPE) of 1.8%. The ensemble model shows an R^2 of 0.76, an MSE of 2.95, and a MAPE of 1.9%. Although the ensemble model shows moderately lower accuracy than the RFs model, considering geological uncertainty, this ensemble model will be very useful for prediction in unknown locations. Future studies should explore the integration of other geotechnical features, as these models relied solely on coordinates. Incorporating features such as alteration patterns along with conventional methods is also suggested for future investigation.

References

[1]. Sinclair, A., & Blackwell, G. (2002). Applied mineral inventory estimation. *UK Cambridge. Univ Press, Cambridge, 400.*

[2]. Rendu, J. (2014). An Introduction to Cut-Off Grade Estimation. *Society for Mining, Metallurgy, and Exploration.*

[3]. Kaplan, U., & Topal, E. (2020). A new ore grade estimation using combine machine learning algorithms. *Minerals, 10, 847.*

[4]. Kapageridis, I., & Denby, B. (1999). Ore Grade Estimation with Modular Neural Network Systems - A Case Study.

[5]. Yamamoto, J. (1999). Quantification of uncertainty in ore-reserve estimation: applications to Chapada copper deposit. *State of Goias, Brazil. Natural Resources Research, 8, 153–163.*

[6]. Wu, X., & Zhou, Y. (1993). Reserve estimation using neural network techniques. *Computer Geoscience, 19, 567–575.*

[7]. Kapageridis, I., & Denby, B. (1998). Neural network modelling of ore grade spatial variability, International Conference on Artificial Neural Networks. *Springer, Bratislava, Slovakia, 209–214.*

[8]. Zimmerman, D., Pavlik, C., Ruggles, A., & Armstrong, M. (1999). An experimental comparison of ordinary and universal kriging and inverse distance weighting. *Mathematics Geology, 31, 375–390.*

[9]. Mahmoudabadi, H., Izadi, M., & Menhaj, M. (2009). A hybrid method for grade estimation using genetic algorithm and neural networks. *Computer Geoscience, 13, 91–101.*

[10]. Jafrasteh, B., Fathianpour, N., & Suárez, A. (2018). Comparison of machine learning methods for copper ore grade estimation. *Computer Geoscience, 22, 1371–1388.*

[11]. Kaplan, U., Dagasan, Y., & Topal, E. (2021). Mineral grade estimation using gradient boosting regression trees. *International Journal of Mining, Reclamation and Environment, 35, 728-742.*

[12]. Bertolini, M., Mezzogori, D., Neroni, M., & Zammori, F. (2021). Machine Learning for industrial applications: A comprehensive literature review. *Expert Systems with Applications, 175.*

[13]. Jung, D., & Choi, Y. (2021). Systematic review of machine learning applications in mining: Exploration, exploitation, and reclamation. *Minerals (Basel, Switzerland), 11, 148.*

[14]. Atalay, F. (2024). Estimation of Fe Grade at an Ore Deposit Using Extreme Gradient Boosting Trees (XGBoost). *Mine Metallurgy Exploration, 41, 2119–2128.*

[15]. Salarian, S., Oskooi, B., Mostafaei, K., & Smirnov, M. (2024). Improving the resource modeling results using auxiliary variables in estimation and simulation methods. *Earth Science Informatics, 17, 4161–4181.*

[16]. Mostafaei, K. & Ramazi, H. (2018). 3D model construction of induced polarization and resistivity data with quantifying uncertainties using geostatistical methods and drilling (Case study: Madan Bozorg, Iran*). Journal of Mining & Environment, 9, 857-872.*

[17]. Mostafaei, K., Maleki, S., Shokri, B., & Yousefi, M. (2023). Predicting gold grade using support vector machine and neural network to generate an evidence layer for 3D prospectivity analysis. *International Journal of Mining and Geo-Engineering, 57, 435-444.*

[18]. Mostafaei, K. & Ramazi, H. (2019). Mineral resource estimation using a combination of drilling and IP-Rs data using statistical and cokriging methods. *Bulletin of the Mineral Research and Exploration, 160, 177-195.*

[19]. Kianoush, P., Afzal, P., Mohammadi, G., Khah, N., & Hosseini, S. (2023). Application of geostatistical and velocity-volume fractal models to determine interval velocity and formation pressures in an oilfield of SW Iran. *Journal of Petroleum Research, 33 (1402-1), 146-170.*

[20]. Sotoudeh F., Ataei M., Kakaie R., & Pourrahimian Y. (2020). Application of Sequential Gaussian Conditional Simulation into Underground Mine Design under Grade Uncertainty. *Journal of Mining and Environment, 11 (3), 695-709.*

[21]. Tahernejad M.M., Khalokakaie R., & Ataei M. (2018). Analyzing the effect of ore grade uncertainty in open pit mine planning; A case study of Rezvan iron mine, Iran. *Int. Journal of Mining and Geo-Engineering, 52 (1), 53-60.*

[22]. Monjezi M., Rajabalizadeh Kashani M. & Ataei M. (2013). A comparative study between sequential Gaussian simulation and kriging method grade modeling in open-pit mining. *Arabian Journal of Geosciences, 6 (1), 123-128.*

[23]. Sohrabian, B., & Tercan, E. (2025). Grade Estimation Through the Gaussian Copulas: A Case Study. *Journal of Mining and Environment, 16(1), 1-13.*

[24]. Parhizkar A., Ataei M., Moarefvand P., & Rasouli V. (2011). Grade Uncertainty and its Impact on Ore Grade Reconciliation between the Resource Model and the Mine. *Archives of Mining Science, 56 (1), 119–134.*

[25]. Parhizkar A., Ataei M., Moarefvand P., & Rasouli V. (2012). A Probabilistic Model to Improve Reconciliation of Estimated and Actual Grade in Open Pit Mining*. Arabian Journal of Geosciences, 5 (6), 1279-1288.*

[26]. Fathi, M., Alimoradi, A., & Hemati Ahooi, H. R. (2021). Optimizing extreme learning machine algorithm using particle swarm optimization to estimate iron ore grade. *Journal of Mining and Environment, 12 (2), 397-411.*

[27]. Ghasemitabar, H., Alimoradi, A., Hemati Ahooi, H., & Fathi, M. (2024). Intelligent Borehole Simulation with python Programming. *Journal of Mining and Environment, 15(2), 707-730.*

[28]. Singh, R., Sarkar, B., & Ray, D. (2021). Geostatistical modelling of a high-grade iron ore deposit. *Journal of the Geological Society of India, 97, 1005–1012.*

[29]. Criminisi, A., Shotton, J., & Konukoglu, E. (2012). Decision forests: a unified approach to classification, regression, density estimation, manifold learning, and semi-supervised learning. *Computer Graphics and Vision, 7.*

[30]. Trehan, S., Carlberg, K., & Durlofsky, L. (2017). Error modeling for surrogates of dynamic systems using machine learning. *International Journal for Numerical Methods in Engineering.*

[31]. Delgado, F., Cernadas, E., Barro, S., & Amorim, D. (2014). Is it necessary to have hundreds

of classifiers to tackle real-world classification challenges? *Journal of Machine Learning Research, 15, 3133-3181.*

[32]. Friedman, J. (2001). Greedy boosting approximation: A gradient boosting machine. *Annals of Statistics, 29, 1189-1232.*

[33]. Batlile, N., Adachi, T., & Kawamura, Y. (2023). Utilization of Artificial Neural Networks in Predicting Copper Ore Grade. *Minerals, 13.*

[34]. Daya Sagar, B. S., Cheng, Q., & Agterberg, F. (Eds.). (2018). Handbook of mathematical geosciences: Fifty years of IAMG. *Springer International Publishing.*

[35]. Griffith, D. (2015). Spatial Statistics and Geostatistics: Basic Concepts. In: Shekhar, S., Xiong, H., Zhou, X. (eds) Encyclopedia of GIS. *Springer, Cham.*

[36]. Rossi, M. E., & Deutsch, C. V. (2014). Mineral Resource Estimation. *New York: Springer.*

[37]. Kaplan, U., & Topal, E. (2020). A new ore grade estimation using combine machine learning algorithms, *Minerals, 10.*

[38]. Atalay, F. (2024). Estimation of fe grade at an ore deposit using extreme gradient boosting trees (XGBoost). *Mining, Metallurgy & Exploration, 41, 2119-2128.*

[39]. Chatterjee, S., Bandopadhyay, S., & Machuca, D. (2010). Ore Grade Prediction Using a Genetic Algorithm and Clustering Based Ensemble Neural Network Model. *Mathematical Geosciences, 42, 309-326.*

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

پیش‌بینی عیار یک معدن سنگ آهن هندی با استفاده از یادگیری ماشین و مدل‌های گروهی

آپان دی[1*]، و گوپینات سامانتا[2]

۱ .محقق، مهندسی معدن، NIT Rourkela، کلکته، هند
۲. تکنسین ارشد، شرکت فولاد تاتا، جمشیدپور، هند

اطلاعات مقاله

تاریخ ارسال: ۲۰۲۵/۰۷/۲۸
تاریخ داوری: ۲۰۲۵/۰۹/۱۴
تاریخ پذیرش: ۲۰۲۵/۱۱/۰۲

DOI: 10.22044/jme.2025.16430.3208

کلمات کلیدی

ML
کریجینگ
XGBoost
جنگل‌های تصادفی
تخمین

چکیده

استخراج مس پورفیری مقادیر قابل‌توجهی باطله تولید می‌کند که به دلیل توانایی در تولید اسید و آزادسازی عناصر بالقوه سمی، مخاطرات جدی زیست‌محیطی و بهداشتی برای انسان به همراه دارد. در این مطالعه، ارزیابی یکپارچه‌ای از ریسک‌های زیست‌محیطی و سلامت انسانی ناشی از باطله‌های معدن مس پورفیری سونگون در شمال‌غرب ایران ارائه شده است. بدین منظور، رویکردی جامع و میان‌رشته‌ای به کار گرفته شد که شامل ترکیب آنالیزهای فیزیکوشیمیایی، کانی‌شناسی و ژئوشیمیایی با روش‌های آماری بود. گونه‌بندی شیمیایی عناصر با استفاده از روش اصلاح‌شده پیشنهادی دفتر مرجع جامعه اروپا انجام شد؛ روشی که در مطالعات متعدد برای ارزیابی تفکیک ژئوشیمیایی و تحرک‌پذیری عناصر به کار رفته است. هدف اصلی این پژوهش، گذار از تحلیل صرف غلظت کل عناصر به سوی ارزیابی دقیق‌تر ریسک مبتنی بر زیست‌دسترس‌پذیری، با بهره‌گیری از چارچوب سازمان حفاظت محیط‌زیست ایالات متحده برای کودکان و بزرگسالان بود. بررسی‌های کانی‌شناسی نشان داد که باطله‌ها دارای پتانسیل خالص تولید اسید هستند، به‌گونه‌ای که مقدار پیریت (حدود ۴ درصد) معمولاً بیش از کانی خنثی‌کننده اصلی، یعنی کلسیت (حدود ۲ درصد)، است. نتایج آنالیزهای ژئوشیمیایی بیانگر غنی‌شدگی قابل‌توجه مس و مولیبدن و همچنین غنی‌شدگی متوسط آرسنیک و کبالت در باطله‌ها بود. در میان عناصر مورد بررسی، بیشترین ضرایب تحرک به ترتیب متعلق به مس (٪۸۱٫۴۹)، سرب (٪۷۶٫۷۱)، روی (٪۷۱٫۶۵) و مولیبدن (٪۵۹٫۲۷) بود. شاخص خطر غیرسرطان‌زایی برای کودکان برابر با ۲٫۰۴ به‌دست آمد که از حد ایمنی فراتر است و در این میان، وانادیم زیست‌دسترس‌پذیر به‌عنوان عامل اصلی ریسک شناسایی شد. این یافته‌ها نشان می‌دهد که اتکای صرف بر غلظت کل عناصر بالقوه سمی می‌تواند گمراه‌کننده باشد و بر ضرورت انجام ارزیابی‌های مبتنی بر گونه‌پذیری شیمیایی برای توصیف دقیق رفتار زیست‌محیطی و مخاطرات سلامت ناشی از باطله‌های معدنی تأکید می‌کند.

نویسنده مسئول مکاتبات: *najmeddinali@ut.ac.ir* (آ. دی)