Journal of Mining and Environment (JME) Vol. 16, No. 3, 2025, 947-962

Shahrood University of Technology

Journal of Mining and Environment (JME)

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

Iranian Society of Mining Engineering (IRSME)

Categorization of Mineral Resources using Random Forest Model in a Copper Deposit in Peru

Marco Antonio Cotrina-Teatino[1]*, Jairo Jhonatan Marquina-Araujo[1], Jose Nestor Mamani-Quispe[2], Solio Marino Arango-Retamozo[1], Johnny Henrry Ccatamayo-Barrios[3], Joe Alexis Gonzalez-Vasquez[4], Teofilo Donaires-Flores[2], and Maxgabriel Alexis Calla-Huayapa[5]

1. Department of Mining Engineering, Faculty of Engineering, National University of Trujillo, Trujillo, Peru
2. Faculty of Chemical Engineering, National University of the Altiplano of Puno, Puno, Peru
3. Department of Mining Engineering, Universidad Nacional San Cristobal de Huamanga, Ayacucho, Peru
4. Department of Industrial Engineering, Faculty of Engineering, National University of Trujillo, Trujillo, Peru
5. Faculty of Industrial Process Engineering, National University of Juliaca, Juliaca, Peru

Article Info

Received 10 January 2025
Received in Revised form 7 March 2025
Accepted 19 March 2025
Published online 19 March 2025

DOI: 10.22044/jme.2025.15568.2984

Keywords

Random forest
Mineral resource categorization
Geostatistics
Kriging variance

Abstract

This work aimed to categorize mineral resources in a copper deposit in Peru, using a machine learning model, integrating the K-prototypes clustering algorithm for initial classification and Random Forest (RF) as a spatial smoother. A total of 318,443 blocks were classified using geostatistical and geometric variables derived from Ordinary Kriging (OK) such as kriging variance, sample distance, number of drillholes, and geological confidence. The model was trained and validated using precision, recall, and F1-score metrics. The results indicated an overall accuracy of 97%, with the measured category achieving 98% precision and an F1-score of 0.98. The total estimated tonnage was 5,859.36 Mt, distributed as follows: 1,446.13 Mt (measured), 2,249.22 Mt (Indicated), and 2,164.01 Mt (Inferred), with average copper grades of 0.43%, 0.33%, and 0.31% Cu, respectively. Compared to the traditional geostatistical methods, this hybrid approach improves classification objectivity, spatial continuity, and reproducibility, minimizing abrupt transitions between categories. The RF model proved to be a robust tool, reducing classification inconsistencies and better capturing geological uncertainty. Future studies should explore hybrid models (K-means with RF, ANN with K-Prototypes, gradient boosting, and deep learning) and incorporate economic variables to optimize decision-making in resource estimation.

1. Introduction

The categorization of mineral resources is a fundamental stage in the mining industry, as it defines the confidence level in the estimation of ore tonnage and grade. This process forms the foundation for strategic planning, mining optimization, and the economic evaluation of mining projects [1–3]. At the international level, the standards established by the Committee for Mineral Reserves International reporting standards (CRIRSCO) classify mineral resources into three main categories: Inferred, indicated, and measured. each category represents different levels of geological certainty and confidence in the continuity of mineralization [4–6]. This classification is of critical importance for mining companies, investors, and financial institutions, as the profitability and feasibility of a project depend on the accuracy of the estimated quantity and quality of mineral resources within a deposit [7–9].

Traditionally, mineral resource classification has been approached using geometric and geostatistical methods, where factors such as drillhole density, spatial continuity of grades, and estimation variance play a crucial role in category assignment [10–12]. Among these approaches, ordinary kriging indicator kriging, and conditional simulation have been widely applied due to their ability to interpolate spatial data and model the uncertainty associated with mineralization [13–15]. However, these methods present certain

Corresponding author: mcotrinat@unitru.edu.pe (M.A. Cotrina-Teatino)

limitations that affect their effectiveness, particularly in structurally complex deposits such as porphyry copper deposits [6, 16]. In particular, these conventional approaches rely on manually defined classification thresholds, introducing subjectivity into the process and reducing the reproducibility of results [17–19]. Additionally, since these methods operate under a block-by-block classification scheme, geostatistical models often generate spatial discontinuities, affecting the geological coherence of the mineral resource model and complicating its application in high-variability scenarios [20, 21].

In recent decades, the development of Machine Learning (ML) techniques has revolutionized numerous scientific and technological fields including mining and geology [22–24]. In this context, ML-based classification algorithms have emerged as promising tools for mineral resource categorization, enabling the automated integration of multiple criteria, eliminating the need for arbitrary threshold definitions, and improving the consistency of models [25, 26]. Recent studies have demonstrated that machine learning and ensemble models such as random forest and StackingC can outperform traditional methods in lithological classification and geochemical anomaly prediction, providing superior results in terms of accuracy and robustness [27, 28]. Similarly, hybrid approaches that combine different ML algorithms with fractal modeling have shown significant improvements in geochemical anomaly detection and mineral resource estimation [28, 29]. Additionally, the use of auxiliary variables in estimation and simulation models has been explored to enhance resource classification and reduce uncertainty in resource modeling, highlighting the importance of integrating multiple data sources in these processes [30]. Among these approaches, Random Forest (RF) has been established as one of the most robust and efficient models due to its ability to handle large volumes of geological data, identify complex spatial patterns in mineralization distribution, and provide interpretability through variable importance analysis.

The application of random forest in mining has been the focus of various recent studies, demonstrating its effectiveness in grade prediction in polymetallic deposits [15], mineral resource classification [31], and hydrothermal alteration identification in epithermal deposits [25]. Unlike traditional methods, RF does not require variogram parameterization or fixed threshold definitions for block categorization, as it automatically learns patterns from the data and generates multiple decision models that optimize the final classification [32]. This not only reduces the inherent subjectivity of geostatistical methods, but also enhances the accuracy and stability of classification in structurally complex scenarios [12, 16].

Despite the growing interest in the use of machine learning models for mineral resource categorization, several challenges remain in their practical implementation. In particular, porphyry copper deposits exhibit highly heterogeneous mineralogical distributions, with abrupt variations in grades, alterations, and geological structures, making the application of conventional categorization models difficult [6, 18]. Therefore, it is essential to evaluate the performance of random forest in these contexts, and determine its effectiveness in comparison to traditional categorization methodologies [16, 19].

The primary objective of this work is to develop and implement a random forest-based model for mineral resource categorization in a copper deposit in Peru. To achieve this objective, the methodology is structured into two main components:

- Mixed multivariable block-by-block clustering: The K-prototypes algorithm will be used to integrate quantitative variables (e.g. grades, distance to drillholes) with qualitative variables (e.g. geological confidence) [33].
- Block smoothing using advanced machine learning techniques: Random Forest (RF) will be applied to improve the spatial coherence of the model and mitigate the discontinuity effect in categorization [18, 20].

This work introduces an innovative hybrid approach that integrates K-prototypes for block classification and random forest for spatial smoothing, enhancing the consistency of mineral resource categorization. Unlike traditional methods, this methodology reduces subjectivity by eliminating fixed thresholds and improving spatial continuity, offering a more objective, reproducible, and adaptable alternative for deposits with high geological variability.

This work is structured into four main sections. Section 2 details the methodology employed, including variable selection, the training, and validation process of the Random Forest (RF) model, and the evaluation criteria. Section 3 presents the results obtained from the case study applied to a copper deposit in Peru, comparing the model's effectiveness with traditional methodologies. Finally, Section 4 discusses the

conclusions of the study and possible lines of future research.

2. Methodology

This research work proposes a machine learning-based approach for mineral resource categorization in a copper deposit in Peru, utilizing the Random Forest (RF) model as the primary tool. The methodology is structured into four fundamental stages: data preprocessing including the selection, transformation, and normalization of variables; model training and optimization, involving hyperparameter configuration, and tuning through cross-validation; validation and performance evaluation, consisting of classification metric analysis and model error assessment; and model application, which focuses on the categorization of mining blocks and spatial coherence analysis.

The implementation was carried out using python 3.12 within the Jupyter notebook interactive environment, employing specialized machine learning and data processing libraries such as scikit-learn [34] NumPy [35], and Pandas [36]. Figure 1 illustrates the methodology for mineral resource categorization. It begins with selecting and standardizing quantitative and qualitative variables, followed by integrating them into a single dataset. The K-prototypes algorithm classifies blocks into measured indicated, and inferred categories. Finally, Random Forest (RF) smoothing enhances spatial coherence, reducing discontinuities and improving classification accuracy.

2.1. Data pre-processing

Data pre-processing is a critical stage to ensure the quality, scalability, and consistency of the data used for mineral resource categorization. This process involved the selection of relevant variables, data transformation, and normalization to optimize the model's performance [37, 38].

Both quantitative and qualitative variables were considered, derived from geostatistical estimation using Ordinary Kriging (OK) in SGeMS [39]. Ordinary Kriging (OK) is a widely used geostatistical interpolation technique for grade estimation, which relies on spatial continuity among sample data points [10, 40]. The system of kriging equations is defined as follows:

$$\sum_{j=1}^{n} \lambda_j \gamma\left(x_i, x_j\right) + \mu = \gamma(x_i, x_0), \quad \forall i = 1, 2, \ldots, n$$
$$\sum_{j=1}^{n} \lambda_j = 1 \tag{1}$$

where $\gamma\left(x_i, x_j\right)$ represents the variogram between data points $x_{i,}$ and x_j. $\gamma(x_i, x_0)$ denotes the variogram between the data points and the estimation point x_0, λ_j are the weights assigned to each sample, and μ is the lagrange parameter, which ensures that the sum of the weights equals one.

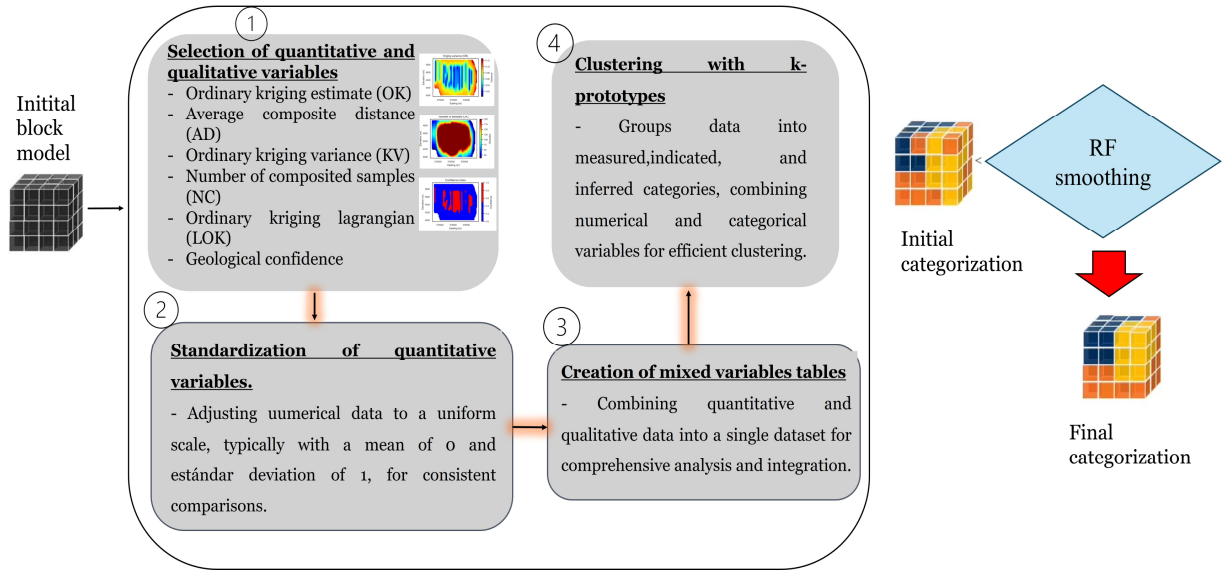

Figure 1. Research workflow.

The fundamental equation of ordinary kriging is:

$$Z^*(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i) \quad (2)$$

where: $Z^*(x_0)$ represents the estimated value at location x_0. $Z(x_i)$ denotes the observed values of the variable at sampling locations. λ_i are the assigned weights, determined through the kriging system of equations. For each block in the model, the following variables were obtained:

- Kriging variance(σ_{OK}^2), which quantifies the uncertainty in grade estimation [13, 14].
- Average distance to samples (d_{OK}), a key geometric indicator of estimation reliability [12].
- Number of drillholes used (n_{OK}), reflecting the density of drillhole data within each block [16].
- Geological confidence, a categorical variable classifying blocks into two categories: high confidence and low confidence [18, 31].
- Kriging lagrangian parameter, which enhances estimation stability. This parameter was employed to correct trends in grade interpolation and improve the spatial consistency of estimated values.

Each block B_j in the model was represented as a multi-dimensional vector defined by:

$$B_j = (x_{1j}, x_{2j}, \dots, x_{Lj}) \quad (3)$$

where x_{Lj} corresponds to each of the L features used in the analysis for block j.

A normalization process was applied to the numerical variables to improve the stability of the model, and reduce biases in the categorization of mineral resources [42, 43]. This transformation ensures that all variables maintain consistent scales, implemented through the power transformer function of scikit-learn [34].

To ensure robust evaluation and prevent overfitting, the dataset was split into 80% for training and 20% for testing. This separation prevents data leakage by ensuring that transformations are learned exclusively from the training set, and correctly applied to the test set, improving the model's generalization capabilities [38, 44].

2.2. Implementation of the random forest model

The random forest model is a supervised learning algorithm based on multiple decision trees, which combines the results of different individual trees to improve accuracy and prediction stability (ensemble learning). This approach is widely used in mining applications due to its ability to handle large volumes of geological data and its robustness against noisy data and outliers [45, 46].

The selection of 200 trees (n_estimators = 200) in the random forest model was based on a sensitivity analysis, where different configurations of trees (ranging from 100 to 200) were evaluated, considering variations in tree depth (max_depth) and splitting criteria. The results (Table 1) indicate that increasing the number of trees slightly improves the model's performance, reaching a maximum test score of 0.903 with 200 trees and a depth of 20, balancing predictive stability and computational efficiency. This justified selecting 200 estimators as the optimal trade-off between accuracy and computational cost. To enhance the reproducibility of the study, grid search cross-validation was employed to optimize hyperparameters. A search space was defined for key parameters such as n_estimators, max_depth, min_samples_split, and min_samples_leaf, applying a 3-fold cross-validation to avoid overfitting. The grid Search results confirmed that 200 trees with a depth of 20 provided optimal accuracy, while preventing unnecessary computational overhead [47, 48].

Based on sensitivity analysis and cross-validation optimization, the optimal hyperparameters for K-prototypes and random forest were selected to ensure maximum model accuracy and stability. The final values are presented in Table 2.

The model was trained using the previously normalized training dataset. Each decision tree in the forest was constructed using bootstrap aggregation (sub-sampling with replacement), which allows the generation of diverse models and reduces variance in predictions. During training, information gain criteria based on the Gini index were applied to optimize node splitting by minimizing impurity at each tree bifurcation [46].

The initial classification of mineral resources into measured, indicated, and inferred was conducted using the K-prototypes algorithm, which integrated key geological and geospatial variables (kriging variance, sample distance, number of drillholes, and geological confidence) to objectively define category boundaries. Once this base classification was established, the Random Forest (RF) model did not reassign categories, but acted as a spatial smoother, reducing discontinuities and enhancing geological coherence. This process mitigated the spotted dog effect, ensuring smoother transitions between

categories and preventing artificial block fragmentation. The relationship with confidence levels was imposed through the variables used in the initial classification and refined by RF, where areas with high geological confidence and dense sampling remained stable, while regions with higher uncertainty reflected in kriging variance and lower sampling density were adjusted through category transitions. Thus, the RF model provided a final classification that was more geologically consistent with the deposit structure.

Table 1. Prueba de sensibilización a diferentes estimadores.

N_estimators	Max_depth	Min_samples_split	Min_samples_leaf	Test_score
100	10	2	1	0.848
200	10	2	1	0.849
100	10	3	1	0.852
200	10	3	1	0.853
100	10	2	2	0.850
200	10	2	2	0.848
100	10	3	2	0.850
200	10	3	2	0.849
100	15	2	1	0.896
200	15	2	1	0.891
100	15	3	1	0.895
200	15	3	1	0.895
100	15	2	2	0.894
200	15	2	2	0.890
100	15	3	2	0.893
200	15	3	2	0.894
100	20	2	1	0.898
200	**20**	**2**	**1**	**0.903**
100	20	3	1	0.895
200	20	3	1	0.899
100	20	2	2	0.891
200	20	2	2	0.894
100	20	3	2	0.890
200	20	3	2	0.891

Table 2. Hyperparameters of the machine learning models used.

Model	Description	Value
K-prototypes	n_clusters	3
	init	"Huang"
	n_jobs	4
	random_state	17276365
	categorical	[3]
Random forest	N_estimators	200
	Max_depth	20
	Random_state	42
	Min_samples_split	2
	Min_samples_leaf	1
	Max_features	log2
	Bootstrap	True

The evaluation of the random forest model was performed on the test dataset using classification metrics commonly applied in data mining [49].

- Accuracy was calculated as the proportion of correct predictions relative to the total predictions made, using the following equation:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total predictions}} \quad (4)$$

- The confusion matrix was employed to assess the model's performance across each category (measured, indicated, and inferred), allowing for the identification of classification biases and improvements in model calibration [50].

The trained random forest model was then applied to the original dataset to assign the corresponding resource category (measured, indicated, or inferred) to each block within the deposit. The results were analyzed using statistical and visual tools to ensure classification coherence and stability [19, 24]. Finally, hyperparameter adjustments and input variable selection refinements were conducted until an optimal level of accuracy and consistency was achieved in mineral resource categorization.

3. Results

3.1. Geological characteristics and case studied area

The open-pit mine studied in this research work is located in the central region of Peru, at an altitude of 4,600 meters above sea level (m.a.s.l.). The geological setting is characterized by a complex lithological composition, consisting of five main rock types: magnetite skarn, granodiorite, dacite

porphyry, calcareous sediments, and volcanic units. The mineralization in the deposit is primarily copper (cu) mineralization, accompanied by molybdenum (mo). The cu-mo mineralization is hosted across the five lithological units, with magnetite skarn (rock 1) and granodiorite (rock 2) being the predominant hosts for high-grade mineralization, while dacite porphyry (rock 3), calcareous sediments (rock 4), and the Catalina volcanic unit (rock 5) show varying degrees of mineralization. The deposit exhibits a porphyry-skarn metallogenic environment, with mineralization occurring in disseminations, veinlets, and massive sulfide replacements, primarily associated with hydrothermal processes and structural controls.

3.2. Exploratory data analysis

The spatial distribution of drillholes within the deposit is shown in Figure 2. Sub-plot (A) illustrates the variability in copper grades (%), while sub-plot (B) displays the distribution of lithological units in the studied area.

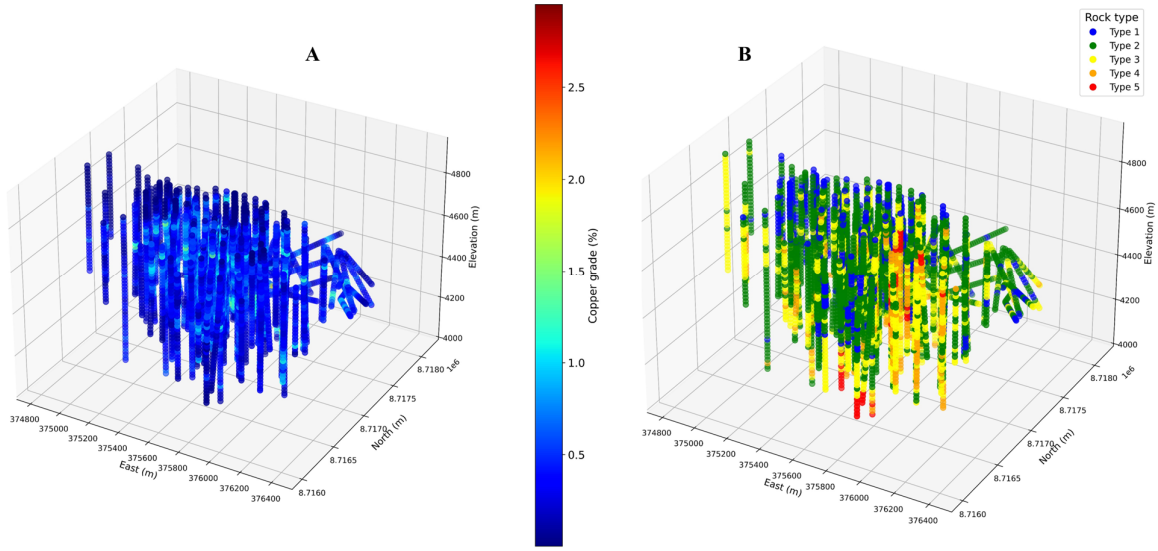

Figure 2. Graphical representation of drillholes. A: copper grade distribution. B: rock type distribution.

A summary of the drillhole dataset is presented in Table 3. The dataset comprises 5,654 drillholes with well-defined spatial coordinates, ranging from 374,821.06 m to 375,824.99 m (east) and 8,716,003.08 m to 8,717,271.73 m (north), with an average elevation of 4,473.54 m. The mean copper grade is 0.43%, with a maximum value of 0.58% and a standard deviation of 0.29%, indicating high mineralization variability within the deposit.

3.3. Block model creation and variables used in categorization

The block model for mineral resource categorization covers a spatial extent of 1,640 m eastward, 2,200 m northward, and 920 m in elevation, defining a significant geological domain. A block size of 20 × 20 × 20 m was applied, resulting in a structured grid with 82 blocks along the East direction, 110 along the North direction, and 46 in elevation. This configuration generates a total of 318,443 discretized blocks, ensuring a detailed and spatially consistent representation of the deposit (see Table 4).

Table 3. Descriptive statistics of the database.

Feature	East (m)	North (m)	Elevation (m)	Copper (%)	Rock type
Count	5,654.00	5,654.00	5,654.00	5,654.00	5,654.00
Mean	375,606.25	8,717,015.68	4,473.54	0.43	2.16
Std dev	307.24	393.54	169.54	0.29	0.78
25%	374,821.06	8,716,003.08	4,050.35	0.00	1.00
50%	375,393.42	8,716,738.40	4,340.07	0.23	2.00
75%	375,602.29	8,716,995.80	4,462.81	0.38	2.00
Maximum	375,824.99	8,717,271.73	4,607.49	0.58	3.00

Table 4. Block model dimensions.

Project coordinates			
Feature	**East (m)**	**North (m)**	**Elevation (m)**
Minimum	374,790	8,715,980	4,015
Maximum	376,430	8,718,180	4,935
Difference	1,640	2,200	920
Block size			
Feature	**East (m)**	**North (m)**	**Elevation (m)**
Size	20	20	20
Number of blocks			
Feature	**East (m)**	**North (m)**	**Elevation (m)**
Blocks	82	110	46

The initial classification was conducted by a Qualified Person (QP), a professional mining engineer, who assigned resource categories (measured, indicated, and inferred) using kriging variance as the primary criterion. This classification served as the baseline for comparison.

The spatial distribution of key variables used in mineral resource categorization reveals significant geological patterns. Ordinary Kriging (OK) estimation highlights heterogeneity in copper grades, with zones of higher mineral concentration. At the deposit edges, the lagrangian parameter of OK displays distinct values, suggesting necessary geostatistical interpolation adjustments. The average distance to samples (AD) and the number of composited samples (NC) indicate a distribution governed by drillhole density. Meanwhile, kriging variance (KV) and the geological confidence index show increased uncertainty in peripheral areas, underscoring their importance in resource categorization (see Figure 3).

A statistical summary of these variables further illustrates the variability in mineralization. The OK estimation results indicate an average copper grade of 0.34% Cu, with a standard deviation of 0.15 and a range from 0.00% to 2.21%, reflecting grade variability. The composite distance (AD) varies between 114.27 m and 433.19 m, with a mean of 231.91 m, indicating differences in drillhole density. The kriging variance (KV) ranges from 0.00 to 0.15, with an average of 0.07, highlighting interpolation uncertainty. The number of composited samples (NC) has a mean of 124, reaching a maximum of 200, ensuring robust data representation. Lastly, the geological confidence index shows an asymmetric distribution, ranging from 0.00 to 1.00, indicating high uncertainty in certain lithological classifications (see Table 5).

3.4. Mineral resource categorization using random forest

The impact of random forest smoothing on mineral resource categorization is evident in the classification results (see Figure 4). The initial classification shows noticeable fragmentation, while RF smoothing enhances spatial continuity, generating a more geologically coherent distribution of measured, indicated, and inferred blocks. Measured resources (red) are mainly concentrated in the central zone, whereas inferred resources (blue) dominate peripheral areas, reflecting lower drillhole density and increased geological uncertainty. This refinement minimizes abrupt transitions, improving the alignment of the classification with the deposit's geological structure.

The confusion matrix evaluates the accuracy of the RF-based mineral resource categorization by comparing the smoothed classification against the initial classifications (see Figure 5). Two comparisons are presented: (A) the classification performed by a Qualified Person (QP) using kriging variance as the primary criterion versus the RF-smoothed categorization, and (B) the initial classification obtained using the K-prototypes algorithm versus the RF-smoothed categorization. The confusion matrix was built by assigning each block to a resource category, and then comparing it to its corresponding classification after RF smoothing. The main diagonal of both matrices highlights the model's ability to correctly classify measured, indicated, and inferred blocks. The highest accuracy is observed in measured resources (98.88% in B, 78.85% in A), followed by indicated (98.44% in B, 83.92% in A) and Inferred (99.61% in B, 96.48% in A). Misclassification errors are mainly concentrated at category transitions, particularly in the QP-based classification, where some Inferred blocks were misclassified as measured, indicating the potential need for adjustments in the categorization thresholds. These

findings align with studies by Nowak *et al.* [12] and Silva *et al.* [16], which highlight increased classification uncertainty in areas with lower drillhole density and higher geological variability.

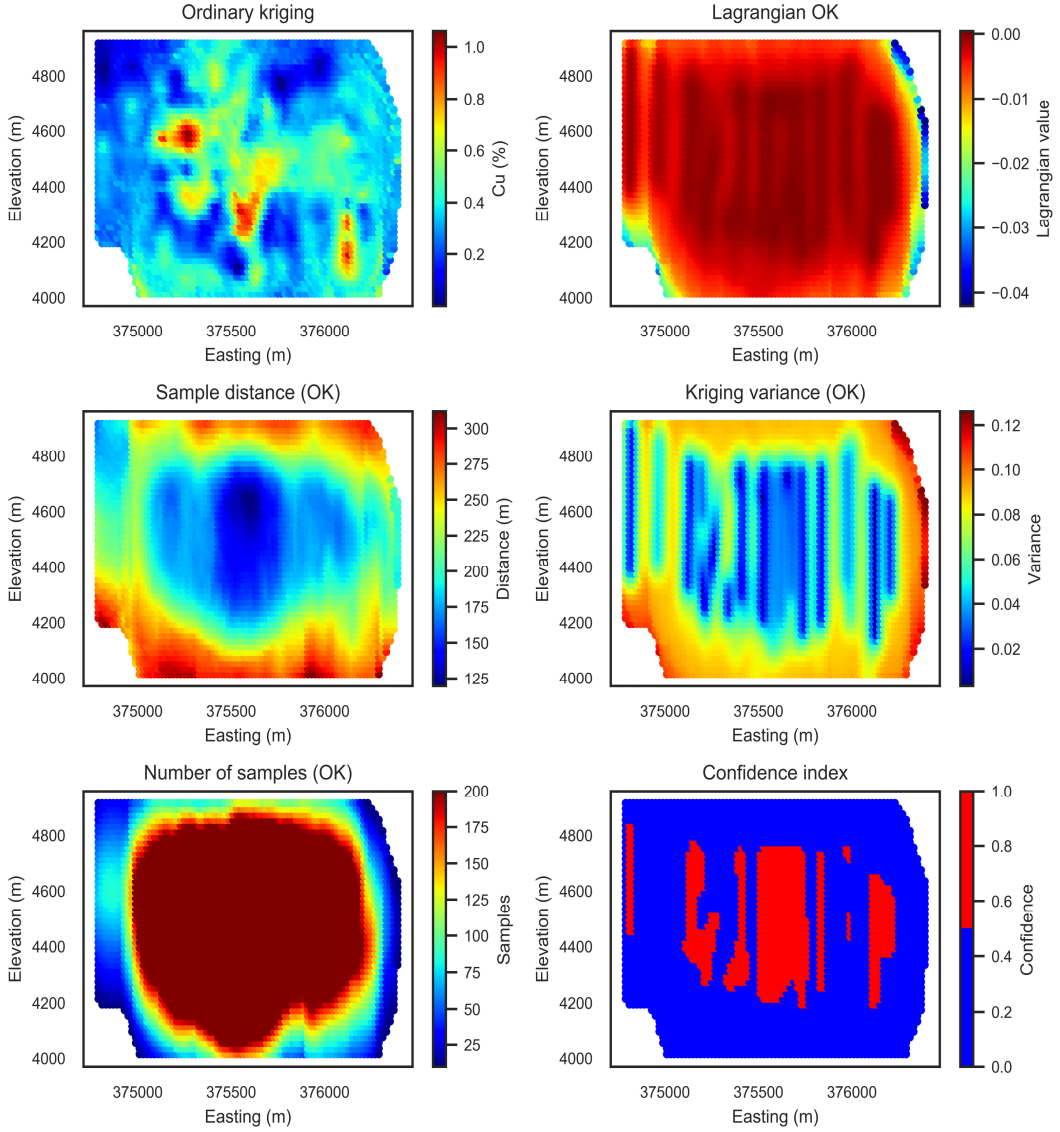

Figure 3. Key-variables used for mineral resource categorization with random forest.

Table 5. Statistical summary of variables used for mineral resource categorization with random forest.

Feature	Ordinary kriging estimate (OK)	Average composite distance (AD)	Ordinary kriging variance (KV)	Number of composited samples (NC)	Ordinay kriging lagrangian (LOK)	Geological confidence
Count	318,443.0	318,443.0	318,443.0	318,443.0	318,443.0	318,443.0
Mean	0.34	231.91	0.07	124	-0.01	0.15
Std dev	0.15	59.19	0.03	70	0.01	0.36
Minimum	0.00	114.27	0.00	10	-0.06	0.00
25%	0.26	189.35	0.05	54	-0.01	0.00
50%	0.33	224.66	0.08	131	-0.00	0.00
75%	0.42	269.69	0.09	200	-0.00	0.00
Maximum	2.21	433.19	0.15	200	0.00	1.00

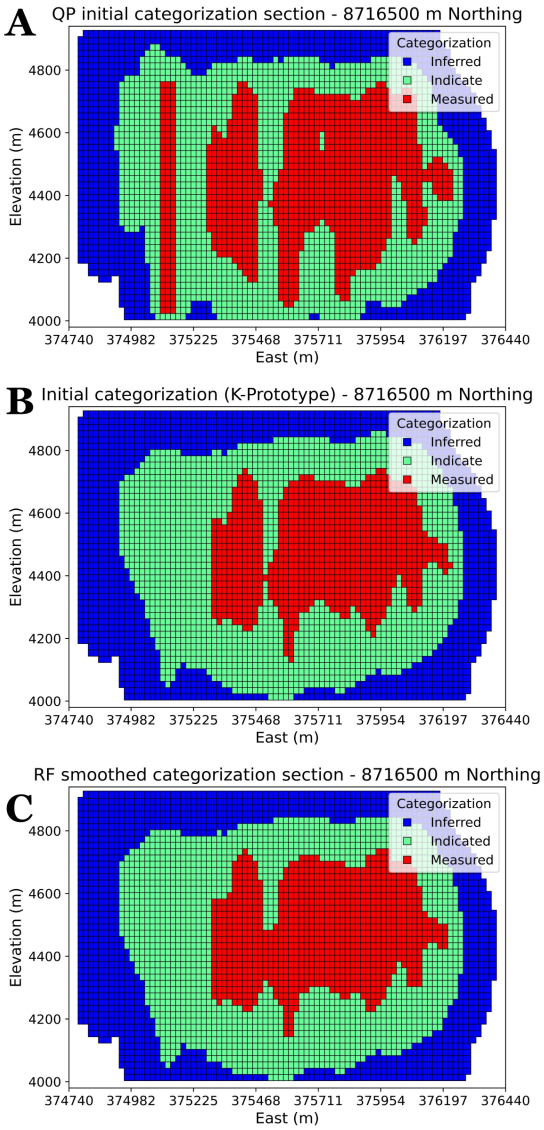

Figure 4. Mineral resource categorization. A: QP initial categorization. B: initial categorization with K-prototype. C: RF smoothed categorization.

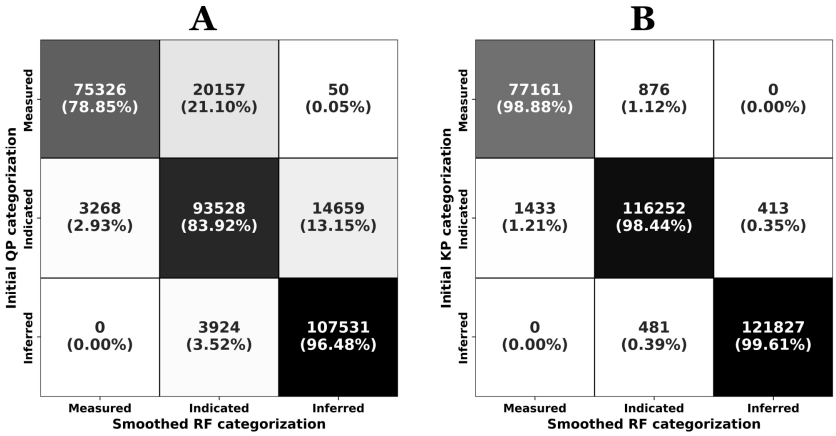

Figure 5. Confusion matrix comparing RF-smoothed categorization against initial categorization. A: comparison between the QP-based categorization and the RF-smoothed categorization. B: comparison between the K-prototype-based categorization and the RF-smoothed categorization.

The performance metrics of the Random Forest (RF) model indicate a high level of classification accuracy, achieving an overall score of 97% (see Table 6). The measured resource category attained the highest precision (96%) and F1-score (0.96), reflecting reliable classification in well-sampled regions. Similarly, the indicated category maintained an F1-score of 0.96, while the inferred category exhibited the highest recall (0.98), suggesting that the model effectively identifies blocks with greater uncertainty while maintaining a low misclassification rate. These results align with findings from Afzal *et al*. [25], who demonstrated that machine learning models excel at capturing complex spatial patterns, leading to a more consistent resource categorization. Additionally, studies by Deutsch *et al*. [10] and Mucha *et al*. [11] emphasize that traditional geostatistical techniques often struggle in heterogeneous deposits, as they require extensive variogram fitting and manual parameter adjustments. In contrast, RF automatically learns spatial relationships, eliminating the need for explicit interpolation models, thereby reducing subjectivity and enhancing reproducibility.

Table 6. Accuracy of mineral resource categorization using the random forest model.

Resources	Precision	Recall	F1-score
Measured	0.96	0.96	0.96
Indicated	0.96	0.96	0.96
Inferred	0.98	0.98	0.98
Accuracy			0.97

The distribution of key variables used for mineral resource categorization highlights differences between the initial QP-based classification and the KP-RF smoothed approach (see Figure 6). The results show that the VOK is generally lower in the KP-RF classification compared to QP, particularly in Indicated and Inferred categories, suggesting that RF smoothing reduces estimation uncertainty and enhances spatial consistency. Similarly, the AD is more evenly distributed in the KP-RF classification, avoiding the extreme values observed in the QP method, which indicates a more geologically coherent classification. The NS also shows an improvement, as KP-RF assigns categories with a more balanced distribution of drilling data, whereas QP presents higher dispersion, particularly in the Indicated and Inferred categories. Finally, the copper grade distribution remains similar between both methods, indicating that the RF smoothing process maintains the geological integrity of the deposit while improving classification consistency.

The confidence level in mineral resource categorization reveals significant variations across different classification regions (see Figure 7). The results show that measured blocks exhibit a higher proportion of high-confidence values (blue), indicating greater reliability in areas with higher drillhole density. In contrast, indicated and inferred blocks predominantly display lower confidence levels (red), suggesting increased uncertainty in regions with sparse sampling. This pattern aligns with the findings of Verly *et al*. [21], who demonstrated that spatial variability plays a crucial role in resource estimation, particularly in large-scale deposits. The improved spatial continuity provided by RF smoothing helps mitigate classification inconsistencies, ensuring a more geologically coherent distribution of confidence levels.

The relationship between tonnage and average copper grade (Cu%) across different resource categories provides insights into the effects of the RF-based smoothing process on mineral classification (see Figure 8, Figure 9 and Figure 10). The tonnage vs. grade curve for measured resources (Figure 8) confirms the expected inverse relationship: as the cutoff grade increases, the available tonnage decreases, while the average copper grade rises. The close agreement between the initial QP classification and the RF-smoothed results suggests that the smoothing process preserves the overall resource distribution without introducing significant distortions.

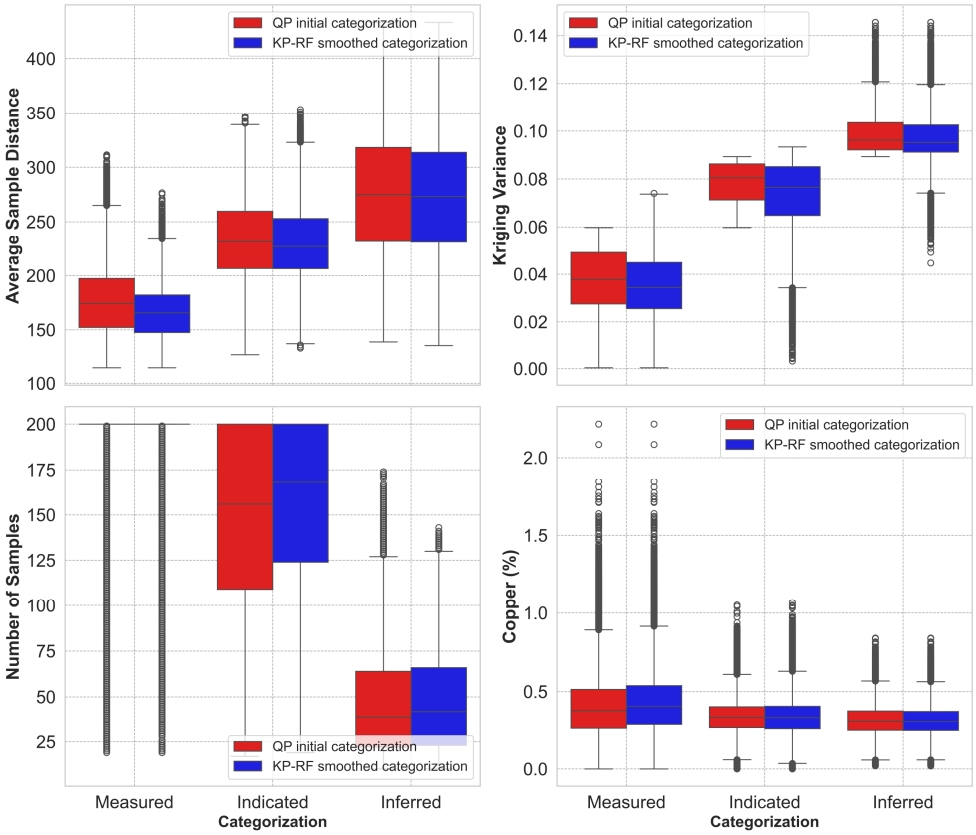

Figure 6. Box-plots of mineral resource categorization comparison.

Similar trends are observed for Indicated resources (Figure 9), where the total tonnage is higher than in the Measured category, reaching approximately 2.2 Mt. However, the lower average copper grade and broader grade dispersion reflect the increased geological uncertainty associated with this classification. The tonnage vs. grade curve for Inferred resources (Figure 10) highlights a greater initial tonnage at lower cutoff grades, with a steeper increase in copper grade as the threshold tightens. This pattern indicates a higher variability in the Inferred category, consistent with its lower drillhole density and greater geological uncertainty. The progressive decline in tonnage with increasing cutoff grade aligns with the findings of Coombes *et al*. [8], reinforcing the reliability of the RF-based approach in capturing mineralization trends, while improving classification consistency.

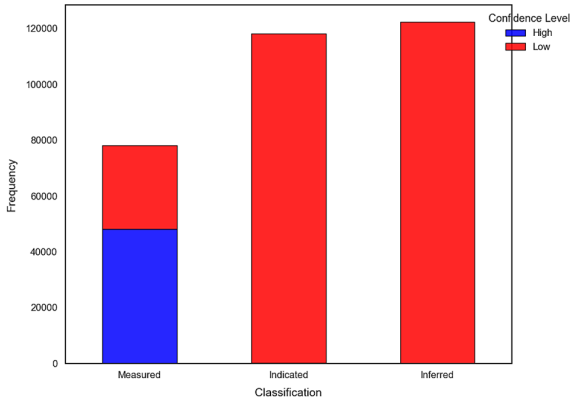

Figure 7. Confidence levels in mineral resource categorization with RF

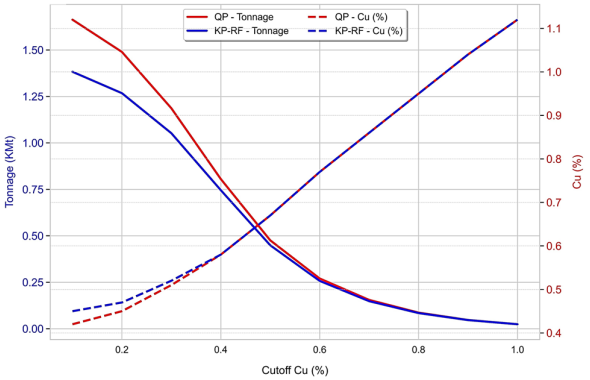

Figure 8. Tonnage vs. average copper grade curve for measured resources using RF.

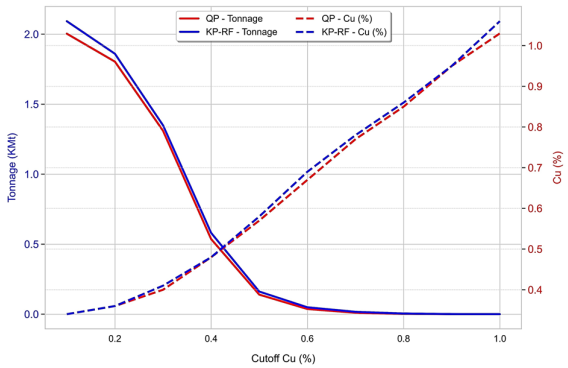

Figure 9. Tonnage vs. average copper grade curve for indicated resources using RF.

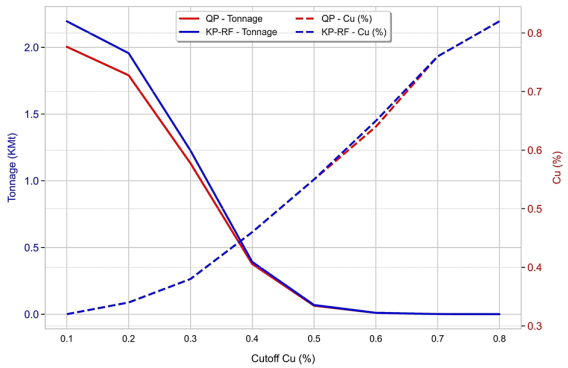

Figure 10. Tonnage vs. average copper grade curve for inferred resources using RF.

The distribution of blocks across mineral resource categories using the RF-based classification approach is shown in Table 7. The Indicated category contains the highest number of blocks (122,240 blocks), followed closely by Inferred resources (117,609 blocks), while measured resources comprise 78,594 blocks. The average copper grade follows a decreasing trend from measured (0.43% cu) to indicated (0.33% cu) and Inferred (0.31% cu) resources. This pattern is consistent with the expected reduction in geological confidence and increasing estimation uncertainty as resources transition from measured to inferred categories.

Table 7. Number of blocks in mineral resource categorization using RF.

Resources	Number of blocks	Copper grade (%)
Measured	78,594.00	0.43
Indicated	122,240.00	0.33
Inferred	117,609.00	0.31

The estimated total tonnage and contained fine copper for each resource category are detailed in Table 8. Measured resources account for 1,446.13 Mt, with 5.60 Mt of contained copper, while Indicated and Inferred resources contribute 2,249.22 Mt and 2,164.01 Mt, containing 6.68 Mt and 6.04 Mt of fine copper, respectively. The Indicated category represents the largest proportion of classified material, followed closely by Inferred resources, suggesting that a significant volume remains in lower-confidence categories.

Table 8. Total material tonnage and contained fine copper in mineral resource categorization

Resources	Material tonnage (Mt)	Contained copper metal (Mt)
Measured	1,446.13	5.60
Indicated	2,249.22	6.68
Inferred	2,164.01	6.04

4. Conclusions

This work introduced an innovative machine learning-based methodology for mineral resource categorization in a copper deposit in Peru, integrating random forest as a spatial smoother to enhance classification coherence and reduce the subjectivity inherent in traditional geostatistical methods. By incorporating geostatistical and geometric variables (kriging variance, sample distance, drillhole density, and geological confidence), the proposed approach improved the spatial consistency of resource classification, while maintaining a high degree of accuracy. The overall classification accuracy reached 97%, with the measured category achieving the highest performance (precision = 0.98, F1-score = 0.98), followed by indicated (precision = 0.96), and Inferred (precision = 0.98), demonstrating the robustness of the RF model in mitigating classification uncertainty.

From a volumetric perspective, the total estimated tonnage of the deposit was 5,859.36 Mt, distributed as follows: 1,446.13 Mt of measured resources, 2,249.22 Mt of indicated resources, and 2,164.01 Mt of Inferred resources. The contained fine copper metal was estimated at 5.60 Mt for

Measured, 6.68 Mt for Indicated, and 6.04 Mt for Inferred resources, highlighting the method's capability to accurately capture mineralization trends and improve resource estimation reliability. The comparative analysis between the traditional QP-based classification, K-prototypes, and the RF-smoothed approach demonstrated that RF significantly reduces category fragmentation, achieving a more geologically consistent classification aligned with the structural complexity of the deposit.

Despite the advantages of RF in improving spatial coherence, some limitations remain. The model's effectiveness may vary in deposits with extreme heterogeneity or strong anisotropy, requiring additional validation in different geological settings. Furthermore, while RF enhances classification consistency, its interpretability is limited compared to rule-based geostatistical methods. For future research, extending this methodology to deposits with higher structural complexity is recommended, incorporating additional economic and operational constraints into the categorization process to improve strategic mine planning. Moreover, exploring hybrid machine learning models, such as combining RF with Gradient Boosting or Deep Learning architectures, could further enhance classification precision. Future work should also integrate uncertainty quantification techniques and conduct an in-depth analysis of variable importance to refine decision-making in mineral resource estimation.

Ultimately, this work provides a novel and reproducible framework for enhancing mineral resource classification through machine learning, offering a more objective, robust, and geologically coherent alternative to conventional geostatistical approaches. The integration of ML-based models in resource estimation represents a promising step toward more data-driven, scalable, and adaptable classification methodologies in the mining industry.

References

[1]. JORC Code. (2012). Australasian code for reporting of exploration results, mineral resources and ore reserves. *AusIMM, 44*.

[2]. CIM. (2019). Estimation of mineral resources & mineral reserves best practice guidelines. *Canadian Institute of Mining*.

[3]. SAMREC. (2016). The South African code for the reporting of exploration results, mineral resources and mineral reserves (the SAMREC Code). *South African Mineral Resource Committee*.

[4]. CRIRSCO. (2014). International Reporting Template for the Public Reporting of Exploration Results Mineral Resources and Mineral Reserves.

[5]. Stephenson, P., Stoker, P. (2001). Mineral resource and ore reserve estimation - the AusIMM guide to good practice (monograph 23). *Miner Eng, 14(9).*

[6]. Abzalov, M. (2016). Methodology of the mineral resource classification. *Modern Approaches in Solid Earth Sciences, 12, 355-363*.

[7]. Owusu, S. (2019). Critical Review of Mineral Resource Classification Techniques in the Gold Mining Industry. *Insights in Mining Science & Technology, 1(3), 555564.*

[8]. Coombes, J., Fahey, G., Stoker, P. (2014). Overview – classifcation and reporting. Mineral resource and ore reserve estimation—the AusIMM guide to good practice. *Australas Inst Min Metall Second Ed Monogr*, *30, 767–770*.

[9]. Menin, R., Diedrich, C., Reuwsaat, JD., De Paula, WF. (2017). Drilling Grid Analysis for Defining Open-Pit and Underground Mineral Resource Classification through Production Data. *Geostatistics Valencia 2016, 19, 271-285*.

[10]. Deustch, C., Leauangthong, O., Ortiz, J. (2007). Case for geometric criteria in resources and reserves classifcation. *Trans Soc Min Metall Explor*, *322*.

[11]. Mucha, J., Wasilewska-Błaszczyk, M., Augušcik, J. (2015). Categorization of mineral resources based upon geostatistical estimation of the continuity of changes of resource parameters. *Proceedings of IAMG 2015 - 17th Annual Conference of the International Association for Mathematical Geosciences*.

[12]. Nowak, M., Leuangthong, O. (2019). Optimal drill hole spacing for resource classification. *Mining Goes Digital - Proceedings of the 39th international symposium on Application of Computers and Operations Research in the Mineral Industry, APCOM 2019*.

[13]. Taghvaeenezhad, M., Shayestehfar, M., Moarefvand, P., Rezaei, A. (2020). Quantifying the criteria for classification of mineral resources and reserves through the estimation of block model uncertainty using geostatistical methods: a case study of Khoshoumi Uranium deposit in Yazd, Iran. *Geosystem Engineering, 23(4), 216-225.*

[14]. Emery, X., Ortiz, JM., Rodríguez, JJ. (2006). Quantifying uncertainty in mineral resources by use of classification schemes and conditional simulations. *Math Geol*, *38, 445-464.*

[15]. Madani, N. (2020). Mineral resource classification based on uncertainty measures in

geological domains. *Springer Series in Geomechanics and Geoengineering, 157-164.*

[16]. Silva, DSF., Boisvert, JB. (2014). Mineral resource classification: A comparison of new and existing techniques. *J South Afr Inst Min Metall, 114(3)*.

[17]. Isatelle, F., Rivoirard, J. (2019). Mineral Resources classification of a nickel laterite deposit: Comparison between conditional simulations and specific areas. *J South Afr Inst Min Metall, 119(10).*

[18]. Rocha, V., Bassani, MA. (2023). Practical application of a multi-layer scorecard workflow (MLSW) for comprehensive mineral resource classification. *Applied Earth Science: Transactions of the Institute of Mining and Metallurgy, 132(3-4)*.

[19]. Dominy, S., Stephenson, P., Annels, A. (2001). Classification and reporting of mineral resources for high-nugget effect gold vein deposits. *Exploration and Mining Geology*, *10, 215-233*.

[20]. Stephenson, PR., Allman, A., Carville, DP., Stoker, PT., Mokos, P., Tyrrell, J., Burrows, T. (2006). Mineral resource classification - It's time to shoot the 'spotted dog'! *Australasian Institute of Mining and Metallurgy Publication Series, 91-96*.

[21]. Verly, G., Parker, HM. (2021). Conditional Simulation for Mineral Resource Classification and Mining Dilution Assessment from the Early 1990s to Now. *Math Geosci*, *53, 279-300.*

[22]. Galetakis, M., Vasileiou, A., Rogdaki, A., Deligiorgis, V., Raka, S. (2022). Estimation of Mineral Resources with Machine Learning Techniques. *Material proceedings, 5(1), 122*.

[23]. Singh, T., Jhariya, DC., Sahu, M., Dewangan, P., Dhekne, PY. (2022). Classifying Minerals using Deep Learning Algorithms. *IOP Conf Ser Earth Environ Sci, 1032(2022), 012046*.

[24]. Hernández, H. (2024). A semiautomatic multi criteria method for mineral resources classification. *Applied Earth Science: Transactions of the Institutions of Mining and Metallurgy, 133, 211-223*.

[25]. Afzal, P., Gholami, H., Madani, N., Yasrebi, A., Sadeghi, B. (2023) Mineral Resource Classification Using Geostatistical and Fractal Simulation in the Masjed Daghi Cu–Mo Porphyry Deposit, NW Iran. *Minerals, 13(3), 370.*

[26]. Battalgazy, N., Madani, N. (2019). Categorization of mineral resources based on different geostatistical simulation algorithms: a case study from an iron ore deposit. *Nat Resour Res, 28, 1329-1351.*

[27]. Farhadi, S., Tatullo, S., Boveiri Konari, M., Afzal, P. (2024). Evaluating StackingC and ensemble models for enhanced lithological classification in geological mapping. *J Geochem Explor, 260, 107441*.

[28]. Afzal, P., Farhadi, S., Konari, MB., Meigoony, MS., Saein, LD. (2022). Geochemical anomaly detection in the Irankuh District using Hybrid Machine learning technique and fractal modeling. *Geopersia*, *12(1), 191-199.*

[29]. Nwaila, GT., Zhang, SE., Bourdeau, JE., Frimmel, HE., Ghorbani, Y. (2024). Spatial Interpolation Using Machine Learning: From Patterns and Regularities to Block Models. *Natural Resources Research, 33, 129-161.*

[30]. Salarian, S., Oskooi, B., Mostafaei, K., Smirnov, MY. (2024). Improving the resource modeling results using auxiliary variables in estimation and simulation methods. *Earth Sci Inform, 17, 4161-4181.*

[31]. Cevik, IS., Leuangthong, O., Caté, A., Ortiz, JM. (2021). On the Use of Machine Learning for Mineral Resource Classification. *Min Metall Explor, 38, 2055-2073.*

[32]. Ribeiro, DT., Filho, CGM., de Souza, LE., Costa, JFCL., de Almeida, D. (2012). Utilização de critérios geoestatísticos para comparação de malha de sondagem visando à maximização da quantidade de recursos. *Revista Escola de Minas, 65(1)*.

[35]. Huang, Z. (1998). Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Min Knowl Discov, 2, 283-304.*

[34]. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., VanderPlast, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12, 2825-2830.*

[35]. van der Walt, S., Colbert, SC., Varoquaux, G. (2011). The NumPy Array: A Structure for Efficient Numerical Computation. *Comput Sci Eng*, *13, 22–30.*

[36]. Álvarez Cid-Fuentes, J., Álvarez, P., Amela, R., Ishii, K., Morizawa, RK., Badia, RM. (2020). Efficient development of high performance data analytics in Python. *Future Generation Computer Systems, 111, 570–581.*

[37]. Ahsan, M., Mahmud, M., Saha, P., Gupta, K., Siddique, Z. (2021). Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance. *Technologies (Basel), 9, 52.*

[38]. Cotrina, M., Marquina, J., Noriega, E., Mamani, J., Ccatamayo, J., Gonzalez, J., Arango, S. (2024). Predicting Open Pit Mine Production using Machine Learning Techniques: A Case Study in Peru. *Journal of Mining and Environment, 15(4), 1345–1355.*

[39]. Wackernagel, H. (2003). Ordinary Kriging. In: Multivariate Geostatistics. *Springer Berlin Heidelberg, Berlin, Heidelberg*, *79–88.*

[40]. Journel, AG. (1983). Nonparametric estimation of spatial distributions. *Journal of the International Association for Mathematical Geology, 15, 445-468.*

[41]. Marquina, J., Cotrina, M., Mamani, J., Noriega, E., Vega, J., Cruz, J. (2024). Copper Ore Grade Prediction using Machine Learning Techniques in a Copper Deposit. *Journal of Mining and Environment, 15(3), 1011-1027.*

[42]. Okwu, MO., Otanocha, OB., Edward, BA., Oreko, BU., Oyekale, J., Oyejide, OJ., Osuji, J., Maware, C., Ezekiel, K., Orikpete, OF. (2024). Investigating the Accuracy of Artificial Neural Network Models in Predicting Surface Roughness in Drilling Processes. *Procedia Comput Sci, 232, 1982-1990.*

[43]. Marquina-Araujo, JJ., Cotrina-Teatino, MA., Cruz-Galvez, JA., Noriega-Vidal, EM., Vega-Gonzalez, JA. (2024). Application of Autoencoders Neural Network and K-Means Clustering for the Definition of Geostatistical Estimation Domains. *Mathematical Modelling of Engineering Problems, 11(5), 1207-1218.*

[44]. Carranza, EJM., Laborte, AG. (2015). Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines). *Comput Geosci, 74, 60-70.*

[45]. Josso, P., Hall, A., Williams, C., Le Bas, T., Lusty, P., Murton, B. (2023). Application of random-forest machine learning algorithm for mineral predictive mapping of Fe-Mn crusts in the World Ocean. *Ore Geol Rev, 162, 105671.*

[46]. Ford, A. (2020). Practical Implementation of Random Forest-Based Mineral Potential Mapping for Porphyry Cu–Au Mineralization in the Eastern Lachlan Orogen, NSW, Australia. *Natural Resources Research, 29, 267-283.*

[47]. Lachaud, A., Adam, M., Mišković, I. (2023). Comparative Study of Random Forest and Support Vector Machine Algorithms in Mineral Prospectivity Mapping with Limited Training Data. *Minerals, 13(8), 1073.*

[48]. Markoulidakis, I., Markoulidakis, G. (2024). Probabilistic Confusion Matrix: A Novel Method for Machine Learning Algorithm Generalized Performance Analysis. *Technologies (Basel), 12(7), 113.*

[49]. Caelen, O. (2017). A Bayesian interpretation of the confusion matrix. *Ann Math Artif Intell, 81, 429-450.*

طبقه‌بندی منابع معدنی با استفاده از مدل تصادفی جنگل در یک کانسار مس در پرو

مارکو آنتونیو کوترینا-تاتینو [1*]، ژایرو جاناتان مارکینا-آراوجو[1]، خوزه نستور مامانی-کوئیسپ[2]، سولیو مارینو آرانگو-رتاموزو[1]، جانی هنری ککاتامایو-باریوس[3]، جو الکسیس گونزالس-واسکوئز[4]، تئوفیلو فلور دونایر، تئوفیلو دونیر الکسی[2] و ماکسگابریل الکسیس کالا-هوایاپا[5]

۱. گروه مهندسی معدن، دانشکده مهندسی، دانشگاه ملی تروخیلو، تروخیلو، پرو
۲. دانشکده مهندسی شیمی، دانشگاه ملی آلتیپلانو پونو، پونو، پرو
۳. گروه مهندسی معدن، دانشگاه ملی سان کریستوبال د هوآمانگا، آیاکوچو، پرو
۴. گروه مهندسی صنایع، دانشکده مهندسی، دانشگاه ملی تروخیلو، تروخیلو، پرو
۵. دانشکده مهندسی فرآیندهای صنعتی، دانشگاه ملی جولیاکا، جولیاکا، پرو

ارسال ۲۰۲۵/۰۱/۱۰، پذیرش ۲۰۲۵/۰۳/۱۹

* نویسنده مسئول مکاتبات: mcotrinat@unitru.edu.pe

چکیده:

هدف این کار طبقه‌بندی منابع معدنی در یک کانسار مس در پرو، با استفاده از یک مدل یادگیری ماشین، ادغام الگوریتم خوشه‌بندی نمونه‌های اولیه K برای طبقه‌بندی اولیه و جنگل تصادفی (RF) به عنوان هموارکننده فضایی بود. در مجموع ۳۱۸۴۴۳ بلوک با استفاده از متغیرهای زمین آماری و هندسی مشتق شده از کریجینگ معمولی (OK) مانند واریانس کریجینگ، فاصله نمونه، تعداد حفره‌ها و اطمینان زمین‌شناسی طبقه‌بندی شدند. این مدل با استفاده از معیارهای دقت، یادآوری و امتیاز F1 آموزش داده شد و اعتبارسنجی شد. نتایج حاکی از دقت کلی ۹۷ درصد بود که دسته اندازه‌گیری شده به دقت ۹۸ درصد و امتیاز F1 0.98 دست یافت. مجموع تناژ تخمینی ۵۸۵۹.۳۶ میلیون تن بود که به شرح زیر توزیع شده است: ۱۴۴۶.۱۳ Mt (اندازه گیری شده)، ۲۲۴۹.۲۲ Mt (نشان داده شده) و ۲۱۶۴.۰۱ Mt (استنباط شده)، با میانگین عیار مس به ترتیب ۰.۴۳٪ و ۰.۳۳٪ ۰.۳۳٪ در مقایسه با روش‌های سنتی زمین‌آمار، این رویکرد ترکیبی عینیت طبقه‌بندی، تداوم فضایی و تکرارپذیری را بهبود می‌بخشد و انتقال ناگهانی بین دسته‌ها را به حداقل می‌رساند. مدل RF یک ابزار قوی است که تناقضات طبقه بندی را کاهش می دهد و عدم قطعیت زمین شناسی را بهتر به تصویر می کشد. مطالعات آینده باید مدل‌های ترکیبی (K-means با RF، ANN با K-Prototypes، تقویت گرادیان، و یادگیری عمیق) را بررسی کرده و متغیرهای اقتصادی را برای بهینه‌سازی تصمیم‌گیری در برآورد منابع ترکیب کنند.

کلمات کلیدی: روش‌های بدون ترانشه، لوله رانی، مدلسازی عددی، تغییر شکل زمین، تحلیل حساسیت.