IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Mathematical Modelling of Engineering Problems
Vol. 11, No. 10, October, 2024, pp. 2625-2638
Journal homepage: http://iieta.org/journals/mmep

Machine Learning Techniques for Predicting the Quantity of ANFO Used in Blasting a Bench in an Open Pit Mine

Marco Antonio Cotrina-Teatino[1*], Jairo Jhonatan Marquina-Araujo[1], Jose Nestor Mamani-Quispe[2], Joe Alexis Gonzalez-Vasquez[3], Solio Marino Arango-Retamozo[1], Eduardo Manuel Noriega-Vidal[1], Teofilo Donaires-Flores[2], Rosa Liliam Quispe-Tello[4]

[1] Department of Mining Engineering, Faculty of Engineering, National University of Trujillo, Trujillo 13001, Peru
[2] Department of Chemical Engineering, Faculty of Chemical Engineering, National University of the Altiplano of Puno, Puno 21001, Peru
[3] Department of Industrial Engineering, Faculty of Engineering, National University of Trujillo, Trujillo 13001, Peru
[4] Department of Mining Engineering, Faculty of Engineering, Technological University of Peru, Lima 04001, Peru

Corresponding Author Email: mcotrinat@unitru.edu.pe

https://doi.org/10.18280/mmep.111004

Received: 24 May 2024
Revised: 16 August 2024
Accepted: 22 August 2024
Available online: 31 October 2024

Keywords:
machine learning techniques, ANFO, blasting, open-pit mining, explosives

ABSTRACT

The objective of this study is to predict the quantity of ANFO required for bench blasting in an open pit mine in Peru, through the application of advanced machine learning techniques. Six models were selected: Artificial Neural Networks (ANN-MLP), Random Forests (RF), Support Vector Machines for Regression (SVR), Extreme Gradient Boosting (XGBoost), K-Nearest Neighbors (KNN), and Bayesian Regression (BR), due to their ability to handle complex multidimensional data and their success in similar applications, such as rock fragmentation prediction. The methodology included the collection of data from 208 drill holes, which were divided into training (70%), validation (15%), and testing (15%) sets. The models were evaluated using RMSE, MSE, MAE, and R2. The KNN model showed the best performance, with an R2 of 0.84, RMSE of 2.37, MSE of 5.60, and MAE of 1.35, standing out in predictive accuracy. This study contributes to the accurate prediction of the ANFO quantity required for bench blasting in open-pit mining, providing a useful tool for improving explosives management based on the specific characteristics of the terrain and operational conditions.

1. INTRODUCTION

The technique of rock fragmentation by blasting is widely recognized for its efficiency and cost-effectiveness in open-pit mines, civil excavations, and tunnel construction [1, 2]. This method relies on using the chemical energy stored in explosives to apply pressure on rock masses, eventually leading to their breakage [3]. In both underground and open-pit mining, the most common explosives are ANFO, emulsion explosives, and dynamite [4]. ANFO, which is a mixture of approximately 94% ammonium nitrate (AN) and 6% fuel oil (FO), is one of the most widely used explosives globally [5, 6]. Furthermore, according to an analysis carried out by S&P Global (Standard and Poor's), the primary purchasers of explosives in industrial settings are China, the United States, the Commonwealth of Independent States, and Central and South America. In addition, a substantial quantity of explosives is utilized in civil engineering endeavors that are not related to the mining sector [7].

Rock fragmentation by blasting is not only effective but also widely used across different sectors. Coal mining is responsible for 40% of total explosives consumption [7]. Moreover, 33% of the world's explosives consumption is attributed to ore mining, and this sector is projected to grow during the forecast period from 2019 to 2024 [7]. Sixteen percent of global explosives consumption is attributed to stone mining for cement production and construction. For instance, in 2020, Biegańska and Barański [8] reported that the usage of emulsion explosives and ANFO in Polish open-pit mining was approximately 7.02 million kg and 17.88 million kg, respectively.

ANFO, as a commonly used explosive is preferred in blasting operations due to several key advantages. Firstly, ANFO is highly cost-effective, being one of the cheapest explosives available on the market, making it an attractive option for large-scale operations that require a significant volume of explosives [1]. Secondly, ANFO is easy to mix and handle, which simplifies its application in the field, especially under mass production conditions [9]. Thirdly, ANFO has a low density, allowing for good gas expansion during blasting, resulting in efficient rock fragmentation [10]. However, it is important to note that ANFO also presents certain challenges and disadvantages. One of the main issues is its lack of water resistance, which limits its use in wet conditions or in underground environments where water presence is common [11]. Additionally, due to its low density, ANFO has lower

energy per unit volume compared to denser explosives, which may require larger quantities to achieve the same effect. Lastly, ANFO is more susceptible to desensitization if not handled properly, particularly in the presence of oils other than fuel oil [12].

Given the widespread use and significance of ANFO in mining, the problem arises in an open pit mine where there are blasting fronts or benches, which are composed of many drilled holes, these holes for blasting need a specific quantity of ANFO (kg), the miscalculation of the quantity of explosive used can lead to blasting with very high fragmentation and more costs in explosive material, affecting the profitability of the company. Some scholars have conducted research on calculations of the number of explosives used for a blast. Chen et al. [13] presented a calculation of charge quantity for blasting shafts, using data from a blasting bench No. 918 of an open pit coal mine called Shengli in Inner Mongolia, used the C++ programming language for their respective calculations, in this mine there were a total of 260 drillings with a total volume of 1244521 m^3, being the final charge quantity of the bench of 214275 kg. Similarly, Duranović et al. [14] indicated that the parameters describing the operation in sublevels are optimal fragmentation and low damage of the rock mass around the stopes, proposed new blasting with a slightly larger quantity, but with a reduced hole diameter, obtained a reduction of explosive use of 174.97 kg where the total charge length was 57.26 m, in a total of 11 boreholes. Likewise, Adhikari et al. [15] designed and executed controlled blasting for plug removal in the same manner for tunnel blasting, the charge per blast round was 60.75 kg, the actual number of perforations varied from 95 to 110. Also, the charge varied according to the face condition by 55 to 65 kg.

In recent years, the application of machine learning (ML) and artificial intelligence (AI) techniques in mining has seen significant advancements. Bui et al. [16] demonstrated that the integration of AI algorithms in the prediction of blast-induced vibrations in open-pit mines can not only enhance operational safety but also significantly reduce costs. Additionally, research by Bakhtavar et al. [17] shows a 23% decrease in operational costs associated with drilling and blasting through the implementation of ML techniques. Bayat et al. [18] conducted in 2020 and 2022, reported reductions in delay costs of 88% and 89%, respectively, by optimizing blasting patterns with ML algorithms [19]. Munagala et al. [20] conducted a comprehensive analysis of current practices in the development of ML to improve drilling and blasting operation. Finally, El Wahab et al. [21] highlighted an increase in the adoption of highly sophisticated hybrid and ensemble ML methods to address key challenges in blasting, with a 72% implementation in surface operations and 38% in underground environments.

Despite the growing body of research on ML applications in mining, it is observed that there is no research that predicts the quantity of ANFO needed for blasting a bench using machine learning. This gap in the literature led to the present study, which aims to predict the quantity of explosive (ANFO) needed to blast rocks in an open pit mine bench using different machine learning methods, such as the Advanced Artificial Neural Networks Multilayer Perceptron Model (ANN-MLP) [22, 23], Random Forests (RF) [24], Support Vector Regression (SVR) [25], Extreme Gradient Boosting (XGBoost) [26], Kernel Nearest Neighbor (KNN) [27], and Bayesian Regression (RB) [28]. These models include crucial spatial data such as east, north, and altitude coordinates, as well as drilled hole height (m), burden (m), spacing (m), stemming (m), booster (kg), and theoretical ANFO (kg). The output variable of the model is the quantity of ANFO (in kilograms) that is needed.

Finally, to guide the reader through the structure of this paper, the essay is organized as follows: section 2 presents an elaborate account of the dataset, the methodology employed, and the data pretreatment procedures carried out. The outcomes generated by the suggested artificial intelligence models are showcased and analyzed in section 3. Section 4 provides a final overview of the study, summarizing the main findings and contributions.

2. MATERIALS AND METHODS

This section presents the dataset, machine learning techniques, and their performance in estimating the quantity of ANFO required for blasting a bench in an open pit mine.

2.1 Data preparation and analysis

The data used in this study were obtained from a blasting bench in an open pit mine located in northern Peru. Data collection was systematically conducted by a specialized technical team, who performed precise measurements of the spatial coordinates of each drill hole (east, north, and elevation) and meticulously recorded the relevant physical characteristics, such as drill hole height, burden, spacing, stemming, booster, and the theoretical amount of ANFO required. This field data collection process was crucial to ensure the accuracy and representativeness of the data, allowing the subsequent analysis to faithfully reflect the actual conditions of the mining operation. Figure 1 shows the graphical representation of the blast bench in its 2D dimensions (East-North).

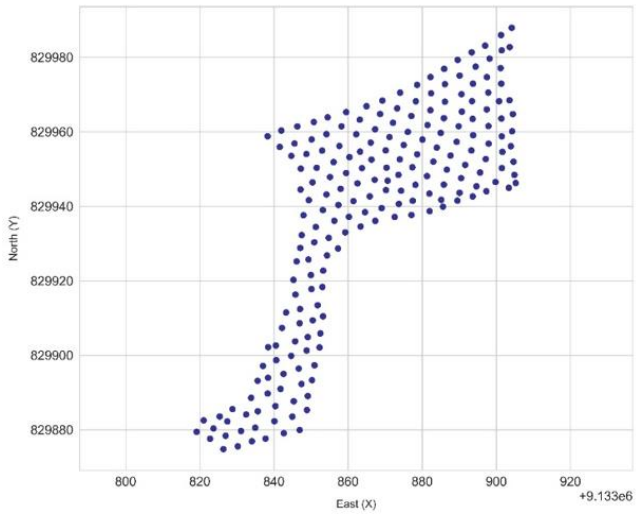

Figure 1. Blasting bench in an open pit mine

The statistical analysis of the data indicated that the mean height of the drilled holes was 8.37 meters, with a variation of 0.13 and a standard deviation of 0.35. Additionally, there are characteristics that remain constant throughout, each with a singular value. These include a burden of 4.10 m, spacing of 4.61 m, stemming of 3.00 m, and booster of 0.45 kg. Table 1 displays the statistical properties of the data.

Table 1. Statistics of the variables used

Variable	Mean	Var	Std	Min	Max
East (m)	9133866.34	524.9	22.91	9133819.1	9133905.24
North (m)	829938.21	906.3	30.11	829874.8	829987.91
Elevation (m)	3347.87	0.13	0.35	3347.1	3348.66
Hole height (m)	8.37	0.13	0.36	7.60	9.20
Burden (m)	4.10	0.00	0.00	4.10	4.10
Spacing (m)	4.61	0.00	0.00	4.61	4.61
Steaming (m)	3.00	0.00	0.00	3.00	3.00
Booster (kg)	0.45	0.00	0.00	0.45	0.45
Theoretical ANFO (kg)	67.30	19.95	4.47	57.00	77.00
Actual ANFO (kg)	64.34	25.18	5.02	54.00	90.00
East (m)	9133866.34	524.9	22.91	9133819.2	9133905.24

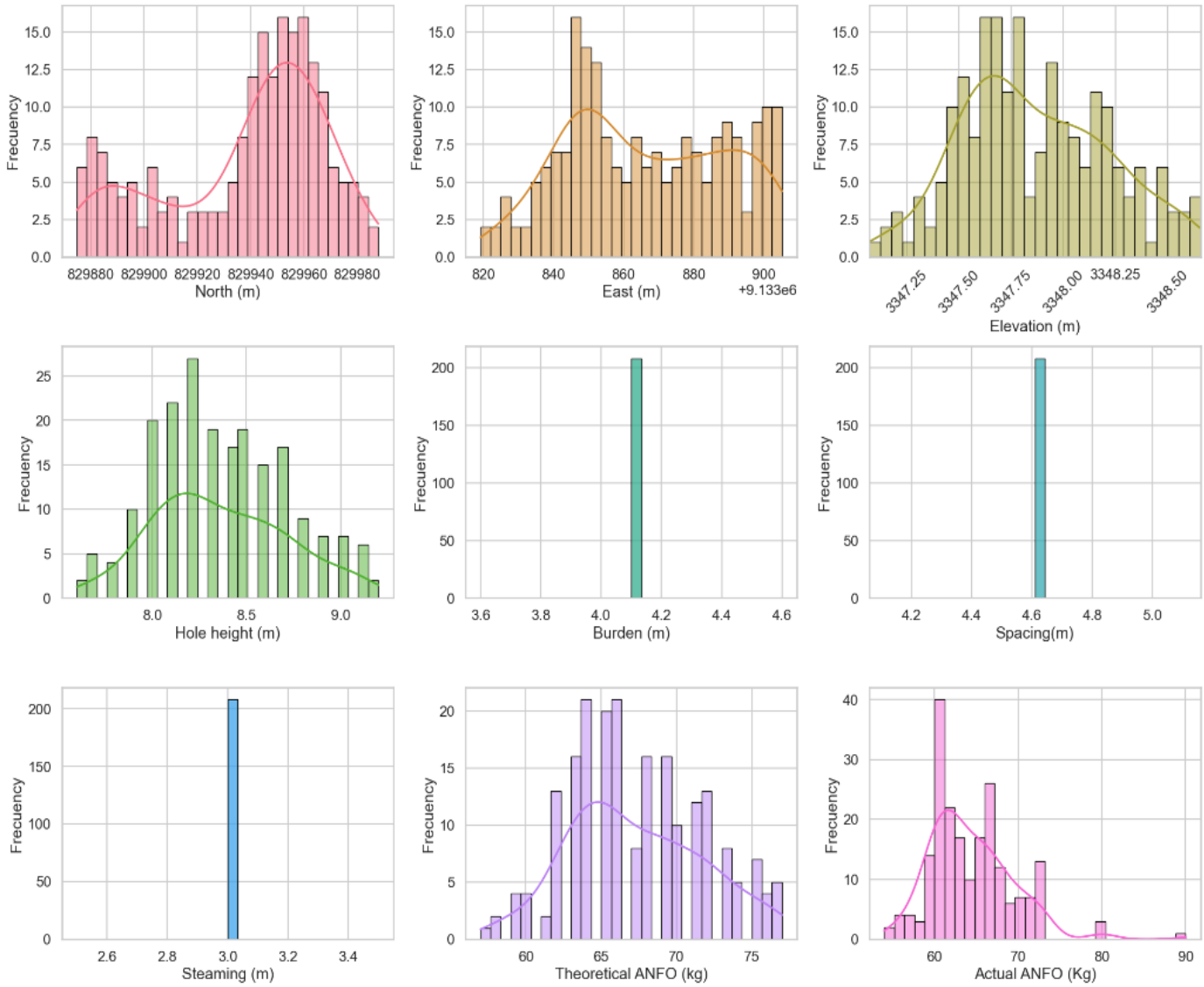

Figure 2. Frequency histograms of the variables used in the prediction

Figure 2 presents the frequency histograms of the variables used in the prediction, showing the variation of drill hole height (m), theoretical ANFO (kg), and actual ANFO (kg), while burden (m), spacing (m), and stemming (m) remain unique and constant values. The distribution of ANFO quantity shows higher amounts in the range of 60 to 70 kg per borehole, and the borehole height is predominantly in the range of 8.0 to 8.5 meters.

Figure 3 illustrates a correlation matrix depicting the linear relationships between geographic and operational variables within the dataset. A coefficient ranging from -1 to 1 is used to quantify correlation. Strong negative linear associations are indicated by coefficients close to -1, and strong positive linear links are indicated by coefficients close to 1. The lack of a linear association is shown by coefficients close to 0. ANFO content (kg) shows a strong correlation with borehole height (m) with a coefficient of 0.78, and with theoretical ANFO, with a coefficient of 0.79. This indicates that the higher the drill height and the theoretical ANFO quantity calculation, the greater the quantity of ANFO required.

During the feature engineering phase, key variables such as drill hole height and theoretical ANFO charge were identified,

showing a strong correlation with the amount of ANFO used. These features were normalized using standardization techniques to ensure equitable contribution to the model. Additionally, feature selection techniques such as Principal Component Analysis (PCA) and Feature Importance were applied, allowing for the reduction of data dimensionality and elimination of redundancies, resulting in significant improvements in model accuracy.

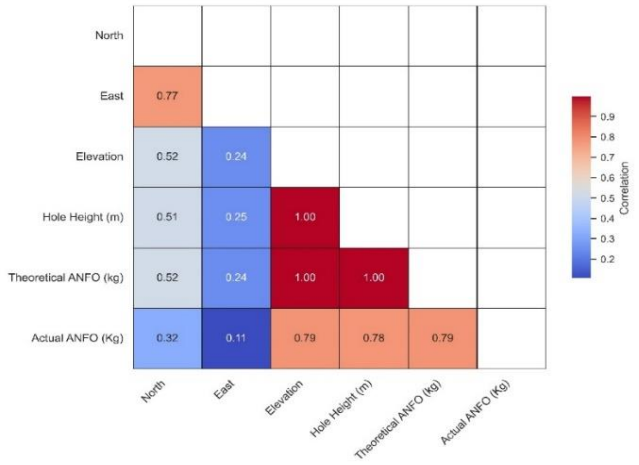

Figure 3. Correlation matrix

To account for the many data types and ranges present in the dataset, including geographic coordinates, bore height (m), burden (m), spacing (m), stemming (m), booster (kg), and theoretical ANFO (kg), the values of each feature were standardized using the mean and standard deviation. The process of normalization was utilized to normalize the values of many attributes to a uniform scale [29]. After dividing by the standard deviation, we subtracted the feature's mean from each data point to regenerate it [30]. Each individual data point, denoted as $x_{i,n}$ was converted into a transformed value, represented as $x'_{i,n}$ using the following procedure:

$$x'_{i,n} = \frac{x'_{i,n} - \mu_i}{\sigma_i} \quad (1)$$

Here, μ and σ represent the average and variability of the i-th characteristic, respectively [31].

Subsequently, the dataset was randomly divided into three subsets: 70% for training, 15% for validation, and 15% for testing. This partitioning was strategically designed to maximize the model's ability to generalize to new data, allowing for robust training, precise tuning during validation, and objective evaluation using independent data. The rationale behind this partitioning lies in ensuring that all critical variables, especially drill hole height and theoretical ANFO charge, were equitably represented in each subset, thus guaranteeing the integrity and representativeness of the modeling process.

Cross-validation was not employed in this study due to the adequate size and representativeness of the dataset, which allowed for a clear division into training, validation, and test subsets. This strategy was sufficient to assess the model's generalization capability, as each subset maintained the distribution of key features. Since the results on the validation set were consistent and reflected the expected performance, cross-validation was deemed unnecessary, which also allowed for greater computational efficiency.

2.2 Artificial Neural Network Multilayer Perceptron (ANN-MLP)

The mining industry has extensively employed artificial neural networks [22]. The multilayer perceptron (MLP) is a neural network that is composed of one or more hidden layers. MLPs have the remarkable ability to approximate any arbitrary function, provided that there is an enough number of nodes in the hidden layers [32]. The computation of the output for a Multi-Layer Perceptron (MLP) with a single hidden layer of H nodes is performed as follows, using an input vector x:

$$\hat{y}(x) = \sum_{j=1}^{H} v_j f\left(w_j^T x + w_{bj}\right) + v_b \quad (2)$$

The weight matrix W = $(w_1, w_2, \ldots, w_H)$ represents the connections between the input layer nodes and the hidden layer in the neural network. The elements of this matrix are denoted as $\{w_{bj}\}_{j=1}^{H}$. The biases of the hidden layer neurons are marked by $\{v_i\}_{j=1}^{H}$. The weights of the connections between the hidden layer and the output node are denoted by $\{v_i\}_{j=1}^{H}$. The output's bias weight is denoted as v_b. The function f is a non-linear activation function. The Rectified Linear Unit (ReLU) function was used as the activation function for the hidden layers in this research.

After evaluating several configurations of artificial neural networks, the most optimal outcomes with the lowest margin of error were obtained by employing a neural network structure with only one input and one output. The configuration consisted of an input layer including 8 neurons, succeeded by four hidden layers consisting of 100, 50, 20, and 10 neurons correspondingly, and an output layer comprising a single neuron (Figure 4). The model was constructed using the TensorFlow 2.15.0 machine learning library and Python version 3.11.7.

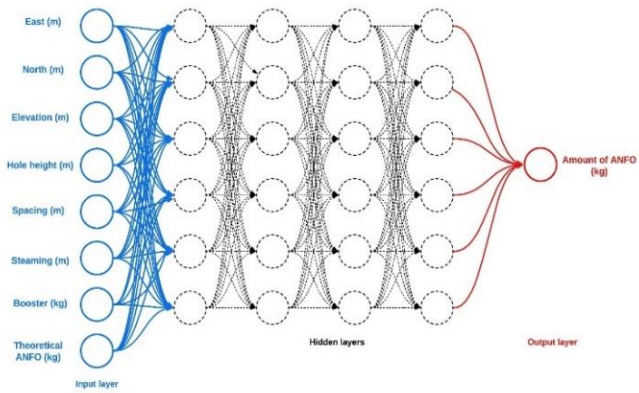

Figure 4. The architecture of ANN-MLP

2.3 Random Forest (RF)

Random Forest (RF) is an ensemble approach that aggregates the results of several random decision trees to formulate a comprehensive prediction. It is applicable to both classification and regression problems of practical interest [33, 34]. Each tree makes binary decisions at its internal nodes based on Boolean tests. For example, if an ordinal attribute is chosen for splitting, the test determines whether the value of that attribute exceeds a given threshold [35]. Instances that

meet the test criteria (i.e., having an attribute value above the threshold) are directed to one child node, while those that do not meet the criteria are directed to the other node. This approach segments the training data into more homogeneous subsets. The purpose of this division is to enhance the differentiation between classes in classification issues or decrease the error in predictions for regression projects. In random trees, the Boolean test at each internal node is picked as the most beneficial split, based on a randomly chosen subset of properties [36]. The growth of the tree persists until further division ceases to enhance the uniformity of the nodes or until a predetermined pruning condition is met, such as a minimum number of occurrences per node or the maximum depth of the tree. Every tree in the forest is built using a Bootstrap sample, which is done independently of the data. This process is comparable to the bagging approach [37].

Using the training examples assigned to each node, the predictions are generated at the tree's terminal nodes, also called leaves. To do this, a sequence of tests is executed beginning at the main node and progressing via intermediate nodes to the appropriate terminal node. To make a prediction in a regression problem, we take the training examples at the leaf node and average their response variables. The final group prediction in classification is determined by a simple majority vote. Regression uses the average of the ensemble's predictions from the random trees to get an ensemble result with a magnitude of T:

$$\hat{y}(x) = \frac{1}{T}\sum_{t=1}^{T}\hat{y}^{(t)}(x) \tag{3}$$

The symbol$\hat{y}^{(t)}(x)$ represents the result produced by the t-th regression tree.

2.4 Support Vector Regression (SVR)

Support Vector Regression (SVR) is a supervised learning technique that relies significantly on the quality and composition of the training and testing datasets to provide accurate results. It is crucial for the model to be effective that both sets have similar distributions [38]. T-statistics are utilized for the analysis of data distribution. The objective of Support Vector Regression (SVR) is to identify a function that minimizes the deviation of all training patterns or datasets from the target values, under a tolerance of ε, while simultaneously maximizing the linearity of the function. Usually, all data points used for training are contained within the range of $(-\varepsilon\ a + \varepsilon)$. The study utilized Support Vector Regression (SVR) and is defined by the equation [39]:

$$K(x, x') = exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right) \tag{4}$$

Sigma (σ) represents the dispersion of the distribution used in the kernel function, whereas $\|x - x'\|^2$ signifies the squared Euclidean distance between two feature vectors.

2.5 Extreme Gradient Boosting (XGBoost)

The described technique refers to a classification or regression method that develops decision trees sequentially. Each generated tree seeks to correct the errors of the previous tree, thus building a sequence of models that gradually minimize the errors and increase the efficiency of the predictions from the beginning of the decision tree algorithm. Such an iterative process continues with the aim of refining the results towards the highest possible accuracy [40].

Considering a data set $D = \{(x_i, y_i)\}$, where x_i belongs to a multidimensional input space R^m and y_i represents the corresponding labels or output values in the real domain R. Within this framework, the tree space is defined such that w_i corresponds to the weight assigned to the i-th leaf of the tree, q symbolizes the tree structure, and T indicates the total number of leaves.

The output function, denoted by $\acute{y}_i$, is modeled as a weighted sum of functions f_k, each belonging to the functional space F.

$$\acute{y}_i = \emptyset(x_i) = \sum_{k=1}^{K} f_k(x_i), f_k \in F \tag{5}$$

where, F is defined as the set of all possible functions of the form.

$$F = \{f(x) = w_{q(x)}\}(q: R^m \to T, w \in R^T) \tag{6}$$

2.6 K-Nearest Neighbors (KNN)

A nonparametric supervised learning classifier, the K-Nearest Neighbor (KNN) or K-NN approach makes predictions or classifications based on how close together data points are. It relies on the clustering of a single data point to determine its categorization [41]. The distance metric used in nearest neighbor methods is the simple Euclidean distance. Specifically, the distance between two patterns $(x_{11}, x_{12}, \ldots, x_{1n})$ and $(x_{21}, x_{22}, \ldots, x_{2n})$ is calculated as follows:

$$DE = \sqrt{\sum_{j=1}^{n}(x_{1j} - x_{2j})^2} \tag{7}$$

2.7 Bayesian Regression

The Bayesian approach, a model for the observed data, provides a vector of unknown parameters usually in the form of conditional density, assuming that it is random and that it has a prior density making use of hyperparameters for its development [42].

2.8 Model evaluation

The Root Mean Square Error (RMSE) would be zero for a model with perfect accuracy. It is computed by taking the square root of the mean squared error between the model's predictions and the observed values [38]:

$$RMSE = \sqrt{\frac{1}{n_{ts}}\sum_{i}(o_i - p_i)^2} \tag{8}$$

To calculate the Mean Absolute Error (MAE), the absolute disparities between the actual and anticipated values are averaged, irrespective of the direction in which they vary. It provides an idea of how large the errors are on average, with a low value being indicative of better model performance [43].

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}| \quad (9)$$

The Mean Squared Error (MSE) is computed by averaging the squared discrepancies between the projected values and the actual values. Unlike RMSE, it does not take the square root at the end [43].

$$MSE = \frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y})^2 \quad (10)$$

The R-squared number, sometimes referred to as the coefficient of determination, measures the amount of variation in the observed and predicted values that can be attributed to the model. An R-squared value above 55% is deemed satisfactory, values below 30% are considered questionable, and values exceeding 75% are regarded as excellent [44]. It is calculated as follows:

$$R^2 = \frac{(\sum(o_i - \bar{o_i})(p_i - \bar{p_i}))^2}{\sum(o_i - \bar{o_i})^2 \sum(p_i - \bar{p_i})^2} \quad (11)$$

2.9 Selection of methods and hyperparameters

The selection of machine learning methods was based on previous studies that have demonstrated their effectiveness in mining applications. Bui et al. [16] used artificial neural networks to predict vibrations in open-pit mines, highlighting the ability of these networks to handle complex nonlinear relationships, which motivated our choice of an ANN-MLP architecture in this study. Similarly, Ohadi et al. [24] validated the use of Random Forests (RF) in predicting blast-induced outcomes, emphasizing their robustness and generalization capability, which aligns with our implementation. The use of Support Vector Regression (SVR) with a kernel has been supported by Mahmoodzadeh et al. [45] who demonstrated its effectiveness in geotechnical predictions of rock fracture toughness. Sun et al. [46] showed that XGBoost performs well in predicting PPV (peak particle velocity) in an open pit mine. The K-Nearest Neighbors (KNN) and Bayesian Regression (BR) models have also been validated in mining contexts, justifying their inclusion in our study [27].

Table 2 presents the hyperparameters associated with each model, optimized to ensure the best possible performance in predicting the quantity of ANFO required.

Table 2. Hyperparameters of machine learning models

Model	Hyperparameters	Value	Description
Multilayer perceptron artificial neural network (ANN-MLP)	Input layers	1	These are the input layers.
	Hidden layers	4	Determines the complexity of the model pattern.
	Neurons in hidden layers	100, 50, 20, 10	Number of neurons per layer, defines how complex the model should be.
	Output layer	1	Prediction by one output layer.
	Activation function	ReLU	Layers' weighted input is transformed using a non-linear function.
	Optimizer	Adam	Works well for improving a neural network's biases and weights.
	Epochs	500	Total number of training times during the process.
	Batch size	100	Number of training sessions in an epoch.
Rando Forest (RF)	Regressor estimators	200	Estimator that fits multiple decision trees.
	Maximum depth	10	Limits the complexity of the tree.
	Minimum samples	4	Minimum samples to form a sheet.
	Minimum division of samples	10	It is the reproductivity of a seed.
	Random state	42	Guarantees results when playing.
	Verbose	0	Provides a lot of details or information about the model.
Support Vector Regression (SVR)	c	1.0	Trade-off between margin and error.
	Core	rbf	Determines the data in SVR.
	Gamma	scale	Defines the form of the function based on the kernel.
	Epsilon	0.1	Determines the tolerance of the loss function.
	Grade	3	Reduce the number of support vectors.
	Tolerance	1e-3	Tolerance Controls the training tolerance.
	Cache size	200	Specifies the memory size during the process.
Extreme Gradient Boosting (XGBoost)	Booster	Gbtree	Uses decision trees as a base.
	Learning rate	0.1	Measures or controls the learning rate.
	Maximum depth	1	Controls the depth of the trees.
	Minimum weight	5	Controls the weight of the sheets.
	N° of estimators	1000	Indicates the number of trees in the set.
	Random state	42	Generate random data process.
	Positive weight	1	Used in a balanced manner.
K-Nearest Neighbor (KNN)	Number of neighbors	5	Defines the number of neighbors.
	Weights	uniform	Weigh neighbors when making a prediction.
	Algorithm	ball_tree	Algorithm for determining nearest neighbors.
	Metric	euclidean	Determines similarities between points based on distance.
	P	2	Parameter detailing a metric.
Bayesian Regression (RB)	N° iterations	300	Number of model iterations to train.
	Tol	1e-3	Tolerance or value that determines the adjustment process (Converged).
	Alfa	1e-6	Regularization applied to the model, generating a penalty to the models.
	Lambda	1e-6	Regularization used for ridge regression type model.
	Intercept Fit	True	True Includes an intercept term in the model fit.

3. RESULTS AND DISCUSSION

Figure 5 depicts the trajectory of the loss function minimization during the training of the Multilayer Perceptron Artificial Neural Network (MLP). The data indicates a decline from initially high values to a consistently lower level as the number of epochs increases. This pattern suggests that the model is a good fit, as the loss function shows a substantial decrease within the initial 50 epochs and then stabilizes around a value of approximately 4.9. This exemplifies the model's capacity to efficiently acquire knowledge from the data without succumbing to overfitting.

3.1 Validation and testing of the machine learning

Figure 6 presents a comparison analysis of anticipated and actual values using the ANN-MLP model. The analysis is partitioned into separate datasets for training, testing, and validation. The training dataset had a correlation value of 0.91, indicating the model's high level of competency in acquiring knowledge from the training data. Similarly, the testing dataset demonstrated a correlation coefficient of 0.84, reflecting a robust correspondence between the model's predictions and the actual ANFO values. The validation dataset showed a correlation value of 0.85, further supporting the model's high capacity to generalize beyond the training data.

Figure 7 illustrates the comparison between the anticipated outcomes and the actual values using the Random Forest (RF) model. The training set demonstrated a correlation coefficient of 0.86, suggesting good accuracy within that dataset. The test and validation datasets showed correlations of 0.89 and 0.96, respectively, indicating excellent predictive performance across different datasets.

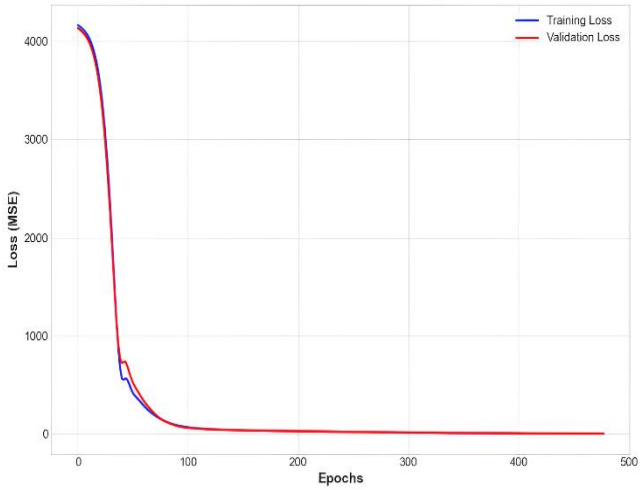

Figure 5. Training and validation curve

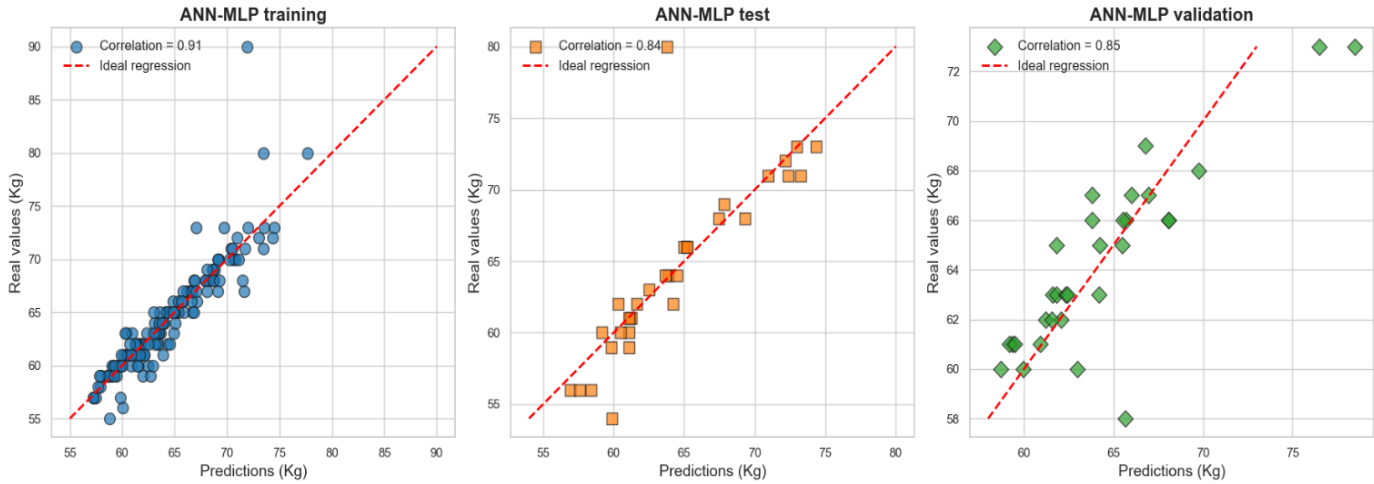

Figure 6. Assessment of expected and actual values with ANN-MLP

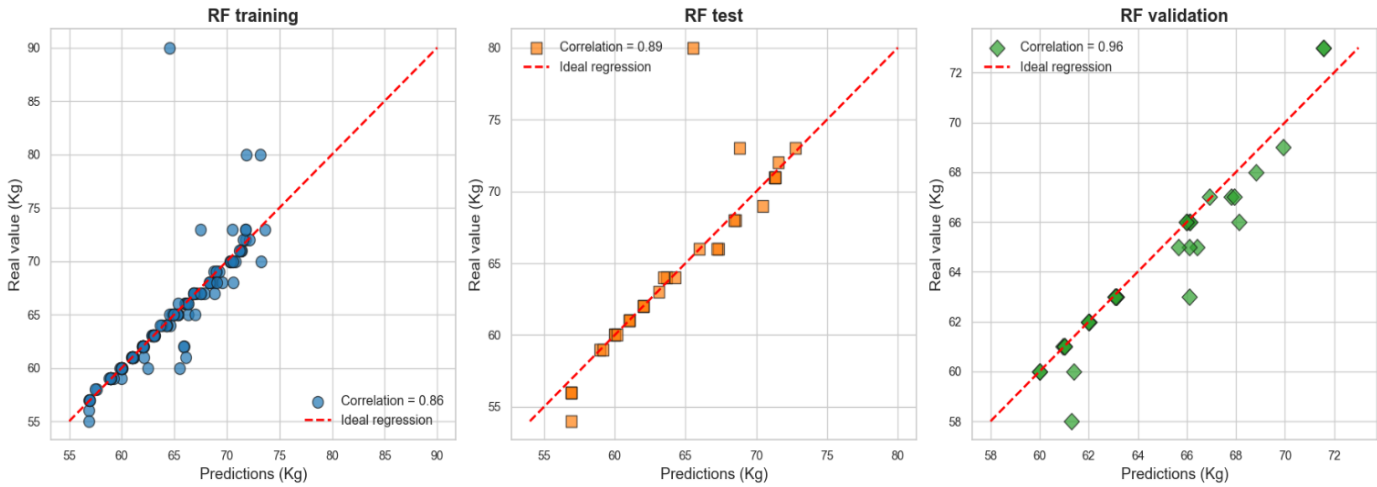

Figure 7. Assessment of expected and actual values with RF

Figure 8 shows the correlation between the values projected by the SVR model and the actual values. The training dataset demonstrated a correlation coefficient of 0.81, indicating significant accuracy. The test and validation datasets exhibited correlation coefficients of 0.74 and 0.98, respectively, reflecting strong predictive capability on the test data and outstanding performance on previously unreported data.

By contrasting the projected values with the actual values in the training, test, and validation datasets, Figure 9 demonstrates the XGBoost model's predictive capability. Correlations of 0.93 on the training set, 0.85 on the test set, and 0.94 on the validation set indicate a robust model capable of accurately predicting the amount of ANFO required for bench blasting.

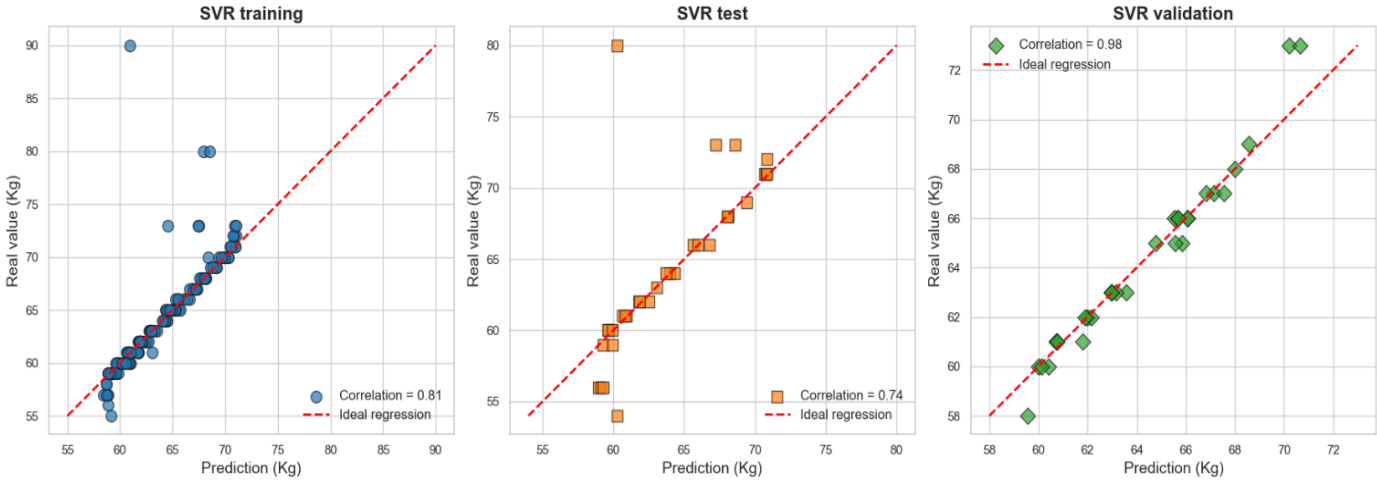

Figure 8. Assessment of expected and actual values with SVR

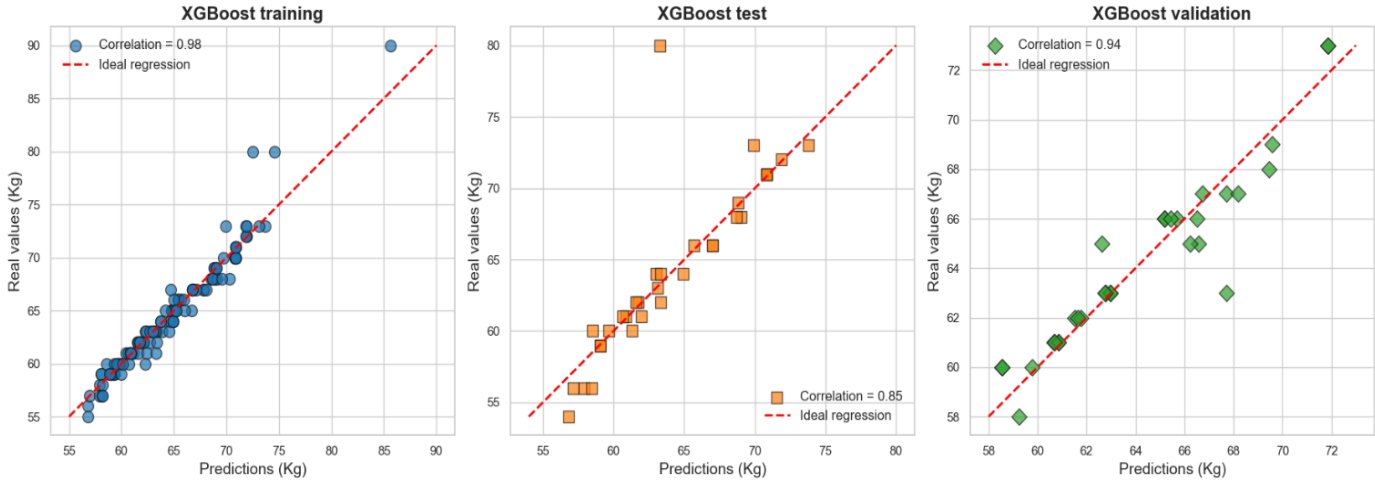

Figure 9. Assessment of expected and actual values with XGBoost

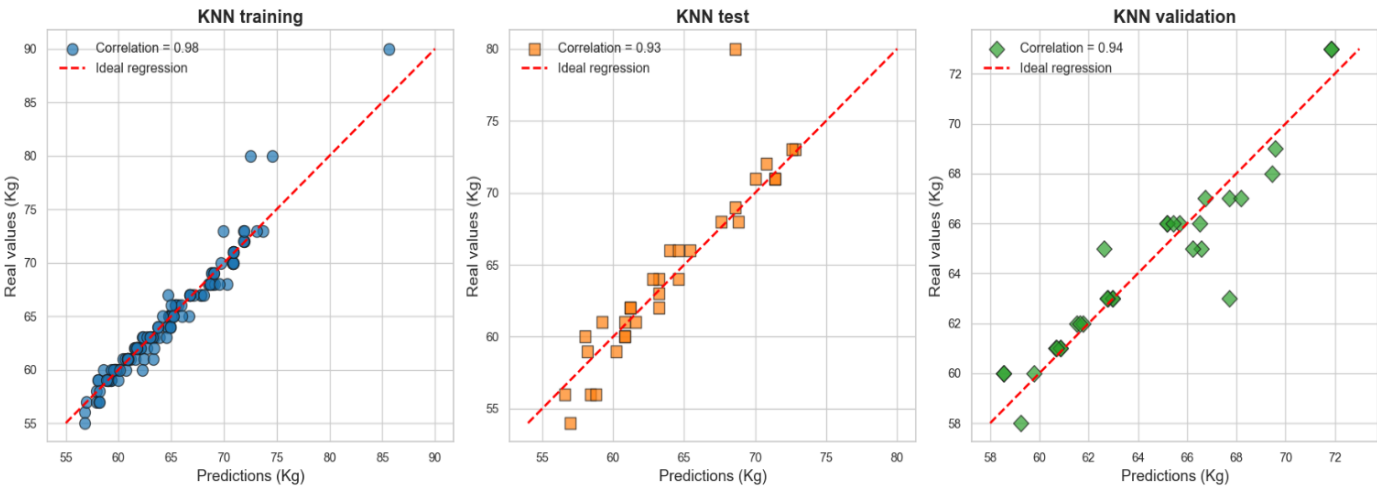

Figure 10. Assessment of expected and actual values with KNN

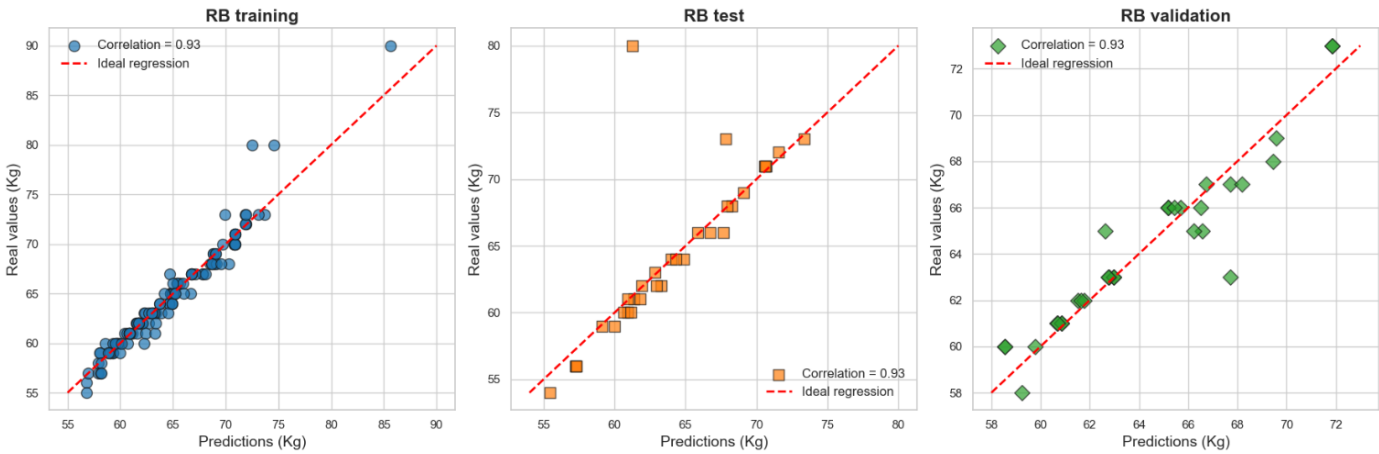

Figure 11. Assessment of expected and actual values with RB

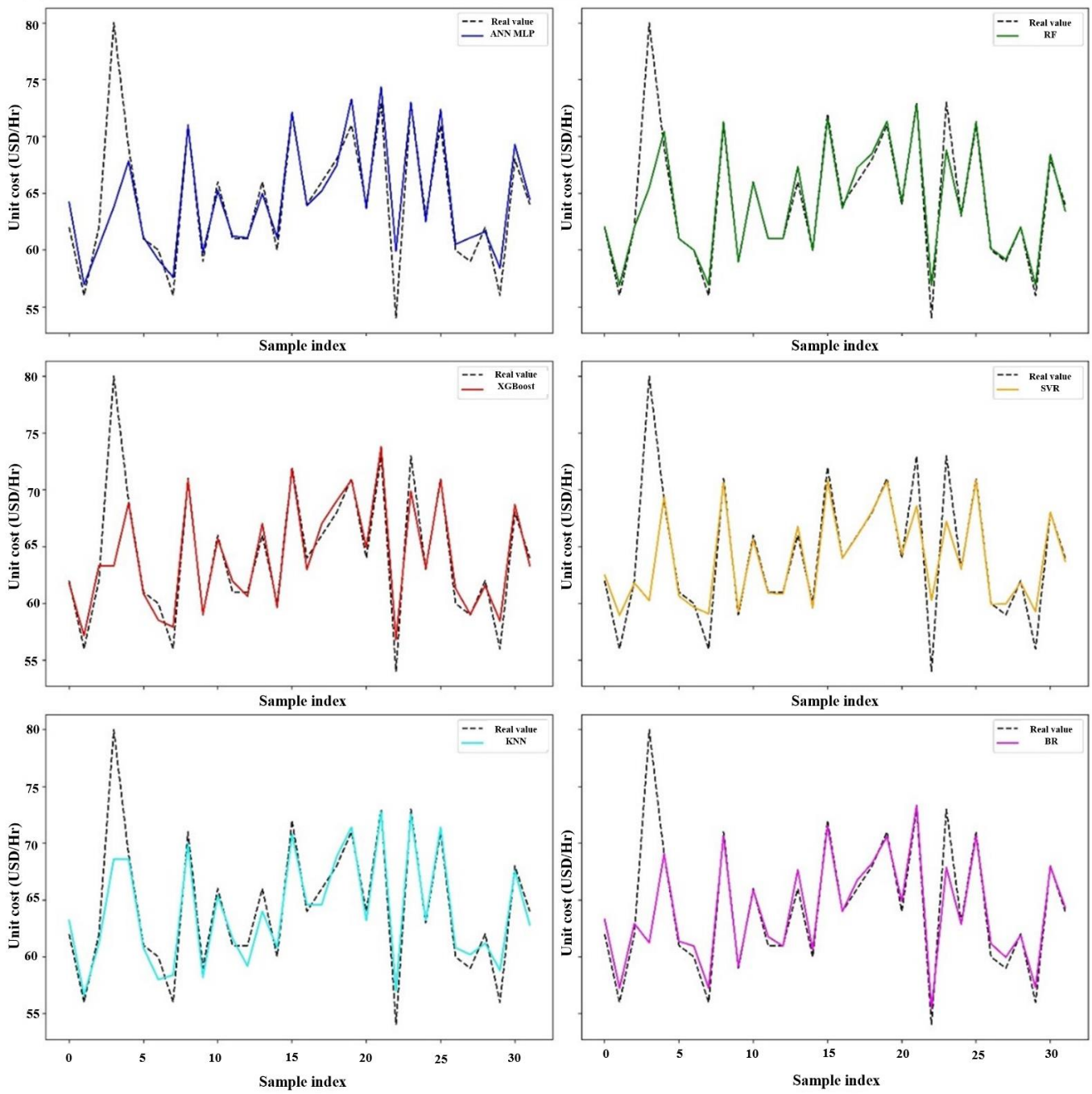

Figure 12. Model prediction results on a test data set

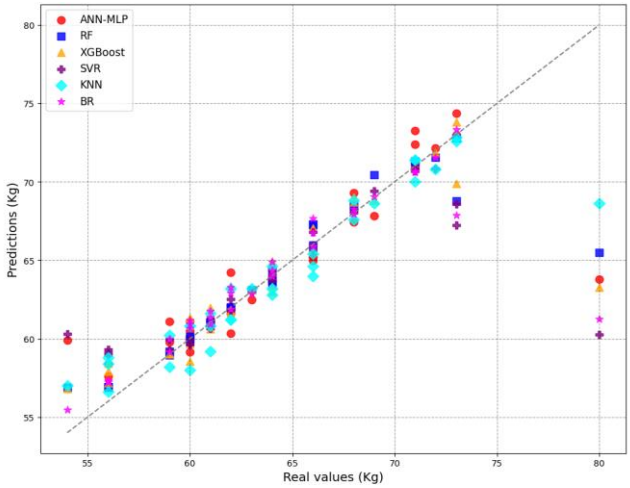

Figure 13. Actual ANFO (kg) vs. predicted ANFO (kg) on the test set for each machine learning model

Figure 10 highlights the K-Nearest Neighbor (KNN) model's prediction accuracy by comparing the actual values with the predicted values across the training, test, and validation datasets. The model accurately predicted the required amount of ANFO for bench blasting, as shown by the correlations of 1.0 for the training set, 0.93 for the test set, and 0.94 for the validation set.

Figure 11 presents the relationship between the predicted and actual values using the Bayesian Regression (BR) model, with datasets broken down into training, test, and validation sets. The model provides accurate predictions for the required amount of ANFO for bench blasting, as the correlations for the training, testing, and validation sets were all 0.93.

3.2 Comparison and evaluation of machine learning models

Figures 12 and 13 depict the results of the machine learning models' predictions on a separate test dataset, facilitating direct comparison. A synopsis of the predicted outcomes is provided in Table 3. The KNN model achieved an R^2 value of 0.84 and an RMSE (Root Mean Square Error) of 2.37, making it the most accurate and reliable model for predicting the amount of ANFO, thereby optimizing efficiency and reducing costs in blasting operations. The Random Forest (RF) model achieved an R^2 value of 0.78 and an RMSE of 2.77. XGBoost's RMSE was 3.17, with an R^2 of 0.72, while the ANN-MLP model had corresponding values of 3.25 and 0.70. The Bayesian Regression (RB) model also performed well, with an R^2 value of 0.65 and an RMSE of 3.52. Finally, the SVR model had the lowest performance with an RMSE of 4.01 and an R^2 of 0.55.

Table 3. Performance of machine learning methods

Metric	ANN-MLP	RF	SVR	XGBoost	KNN	RB
RMSE	3.25	2.77	4.01	3.17	2.37	3.52
MSE	10.55	7.68	16.12	10.06	5.60	12.41
MAE	1.85	1.02	1.66	1.35	1.35	1.33
R^2	0.70	0.78	0.55	0.72	0.84	0.65

The performance metrics such as RMSE, MSE, MAE, and R^2 have significant practical implications in operational environments. KNN, with the lowest RMSE (2.37) and a high R^2 (0.84), stands out as the most accurate and reliable model for predicting the amount of ANFO, optimizing efficiency, and reducing costs in blasting operations. The low MAE of RF (1.02) suggests consistent predictions, while a low MSE in KNN (5.60) reinforces its ability to minimize severe errors, directly impacting operational safety and effectiveness.

Table 4. Confidence intervals for error metrics

Model	Mean RMSE	Confidence Interval (95%)
ANN-MLP	2.97	[1.04 – 5.32]
RF	2.47	[0.58 – 4.64]
SVR	3.70	[1.35 – 6.51]
XGBoost	2.88	[0.91 – 5.28]
KNN	2.21	[1.01 – 3.75]
RB	3.09	[0.71 – 5.88]

Table 4 presents the 95% confidence intervals for the mean RMSE of each model. These intervals reflect the variability in prediction error and provide a measure of each model's precision. The results indicate that the K-Nearest Neighbors (KNN) model achieved the lowest mean RMSE (2.21) with a relatively narrow confidence interval [1.01 – 3.75], suggesting consistent and reliable performance. Conversely, the SVR model exhibited the highest mean RMSE (3.70) and the widest confidence interval [1.35 – 6.51], indicating greater variability and lower accuracy in its predictions.

Table 5. Student's t-test for model comparison

Model Comparison	t-Statistic	p-Value	Significance
RF vs. SVR	-19.59	$2.60e^{-78}$	Significant
RF vs. XGBoost	-6.89	$7.62e^{-12}$	Significant
RF vs. KNN	5.51	$4.07e^{-08}$	Significant
RF vs. RB	-9.37	$1.86e^{-20}$	Significant
RF vs. ANN	-8.85	$1.91e^{-18}$	Significant
SVR vs. XGBoost	12.28	$1.81e^{-33}$	Significant
SVR vs. KNN	26.87	$5.49e^{-136}$	Significant
SVR vs. RB	8.55	$2.51e^{-17}$	Significant
SVR vs. ANN	11.29	$1.06e^{-28}$	Significant
XGboost vs. KNN	12.89	$1.36e^{-36}$	Significant
XGBoost vs. RB	-2.93	0.0034	Significant
XGBoost vs. ANN	-1.52	0.128	Not significant
KNN vs. RB	-14.91	$9.30e^{-48}$	Significant
KNN vs. ANN	-15.62	$5.44e^{-52}$	Significant
RB vs. ANN	1.64	0.101	Not significant

Table 5 presents the results of the student's t-test, used to assess whether the differences in performance between pairs of models are statistically significant. P-values less than 0.05 indicate that the observed differences are not due to chance and are therefore significant. The comparison between RF and SVR revealed a highly significant difference (t=-19.59, $p=2.60e^{-78}$), confirming that RF has significantly better performance than SVR. Similarly, significant differences were observed in comparisons between RF and XGBoost, RF and RB, and RF and ANN-MLP. Comparisons between SVR and KNN (t=26.87, $p=5.49e^{-136}$) and SVR and RB (t=8.55, $p=2.51e^{-17}$) were also highly significant, indicating that KNN and RB outperform SVR in terms of accuracy.

Based on the results obtained, KNN stands out as the optimal model for predicting the amount of ANFO, as shown by a comprehensive analysis of performance metrics, confidence intervals, and statistical tests. KNN achieved the lowest RMSE (2.37) and MSE (5.60), along with the highest R^2 (0.84), demonstrating its high accuracy and predictive

capability. Additionally, its confidence interval for RMSE [1.01 – 3.75] is relatively narrow, indicating consistency in predictions. The student's t-tests also support this choice, showing statistically significant differences between KNN and other models, thus consolidating KNN as the best option for this task.

3.3 Data sensitization test

Figure 14 shows the results of a sensitivity analysis for ANFO predicted using the KNN model, focusing on two independent variables: borehole height (m) and theoretical ANFO. In the first plot (left), corresponding to borehole height (m), it is observed that the quantity of ANFO remains almost constant around 63.95 kg for borehole heights up to 10 meters, but experiences a significant increase to 64.01 kg when the borehole height reaches 12 meters. This suggests a threshold beyond which more ANFO is required to maintain blasting efficiency. In contrast, the relationship between theoretical ANFO and the quantity of ANFO used shows a clearly linear and ascending pattern, where the increase is almost proportional, starting at 62.11 kg ANFO for 50 kg theoretical ANFO and increasing to 65.36 kg ANFO for 74 kg theoretical ANFO. This trend indicates a strong positive correlation between the calculated theoretical ANFO and the actual ANFO quantity. These variables have a significant influence on the prediction of the quantity of ANFO needed, as increasing the drill hole height significantly increases the required ANFO.

3.4 Quantity of ANFO required for bench blasting

Figure 15 presents a comparative analysis of various machine learning models regarding the total amount of ANFO required for bench blasting and their respective coefficients of determination, which assess how well the models fit the data. It is observed that the models differ both in the predicted quantity of ANFO and in their accuracy of fit. With an R^2 of 0.78, the Random Forest (RF) model shows an excellent match to the observed data and predicts the largest amount of ANFO required, with a value of 13,385.87 kg. In contrast, the Support Vector Regression (SVR) model estimates the lowest quantity of ANFO required, at 13,307.16 kg, but with an R^2 of 0.55, suggesting a less accurate fit. Among the models compared, the K-Nearest Neighbors (KNN) model stands out with the highest R^2 of 0.84, indicating excellent predictive capability, and it predicts an ANFO requirement of 13,322.6 kg.

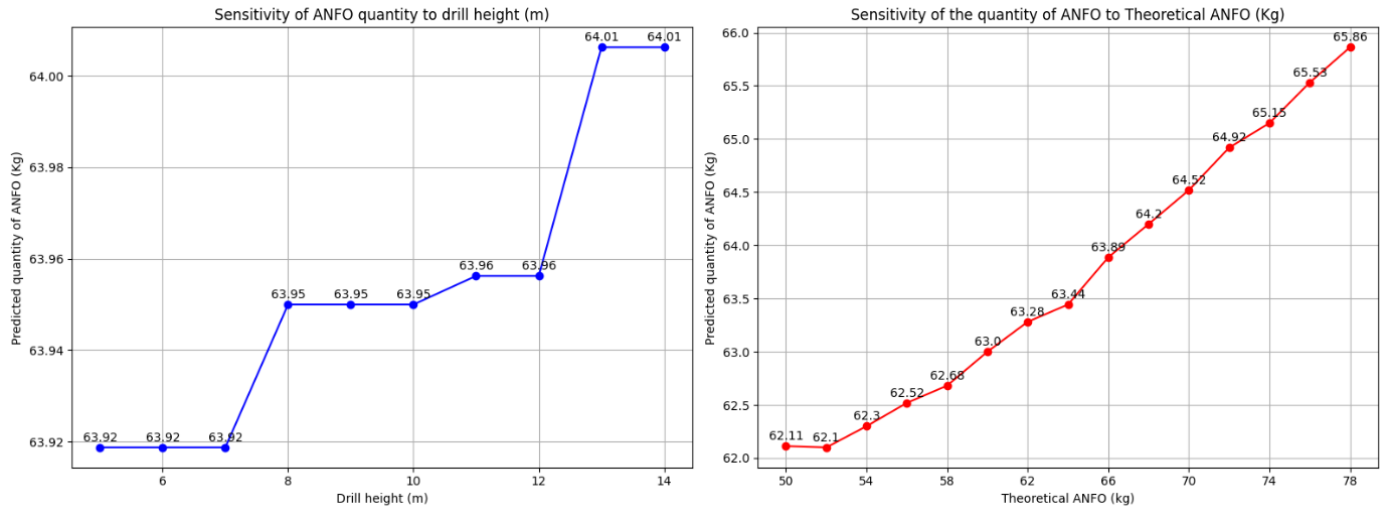

Figure 14. Sensitization test of the quantity of ANFO used

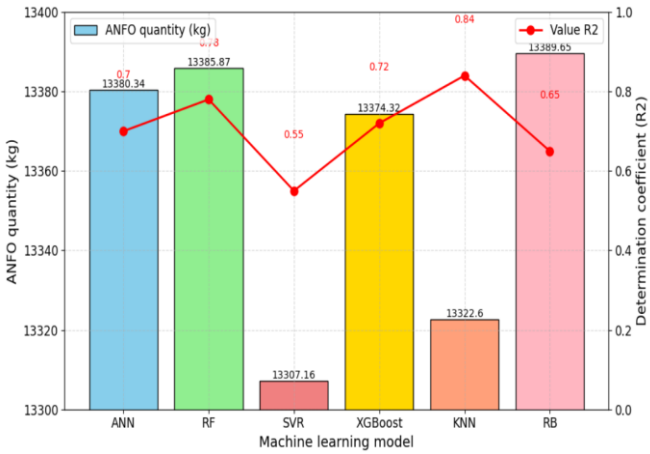

Figure 15. Quantity of ANFO required for bench blasting

Taking into consideration that the KNN model is the most accurate, the total quantity of ANFO required to blast the bench with 208 drilled holes is 13,322.6 kg. In comparison, Chen et al. [13] required a total of 214,275 kg of explosives for blasting their study bench (260 holes), with a total volume of 1,244,521 m^3. There is a significant difference in the quantity of explosives used, but Chen et al.'s [13] study involved various types of explosives, not just ANFO. Similarly, Duranović et al. [14] used a total of 174.97 kg of explosives for optimal fragmentation in a total charge length of 57.26 m, with only 11 drilled holes in their study. Adhikari [2], in his research on controlled blasting in tunnels, used 60.75 kg of explosives, with the number of drilled holes varying from 95 to 110, and the charge quantity varied according to the face condition, ranging from 55 to 65 kg.

3.5 Study limitations

Each of the evaluated models presents limitations that must be considered in practical applications. KNN, while demonstrating excellent overall performance, can be affected by the quality and size of the dataset, as its performance

declines in the presence of large volumes of noisy or unbalanced data. RF, on the other hand, may encounter overfitting issues when too many trees are applied, especially if hyperparameters are not properly tuned. SVR is sensitive to the selection of parameters such as C and gamma, and its performance may degrade if these are not optimized for the specific data. XGBoost, though powerful, can be prone to overfitting if model complexity is not controlled. RB may struggle in contexts with high collinearity among the predictor variables. Finally, ANN-MLP requires careful tuning of the architecture and a large dataset to avoid convergence issues and overfitting, making it less suitable for small or noisy datasets.

3.6 Comparison with existing methods

Recent literature has seen the development of several innovative approaches to improve the accuracy of predicting explosive use in mining operations. For example, Dennis, and Rigby [47] proposed the Direction-Encoded Neural Network (DeNN), an artificial neural network designed to quickly predict explosive charges in obstructed scenarios. This approach achieved an R^2 of 0.97, demonstrating high accuracy in predicting blast pressure using 1 kg TNT charges. On the other hand, Majid et al. [48] utilized the hybrid RUN-XGBoost model, which combines the XGBoost algorithm with Runge-Kutta (RUN) optimization to enhance parameter selection and the accuracy of predicting peak particle velocity (PPV) induced by blast vibrations in an open-pit mine. This model achieved an R^2 of 0.96, showcasing its effectiveness in the blasting context.

When comparing the results of this research with the study conducted by Mishra et al. [3], it is evident that Mishra achieved superior performance with the XGBoost model, obtaining an R^2 of 0.978, MAE of 0.779, and RMSE of 1.066. In contrast, the XGBoost model in this study obtained R^2=0.72, MAE=1.35, and RMSE=3.17. Similarly, Mishra obtained R^2=0.874, MAE=1.528, and RMSE=2.543 for the KNN model. Compared to Mishra's model, the suggested KNN model in this study performed better in predicting the amount of ANFO (kg), with an R^2 value of 0.84, MAE value of 1.35, and RMSE value of 2.37.

In comparison, the models developed in the present study, such as KNN, which obtained an R^2 of 0.84, and RF, with an R^2 of 0.78, demonstrate competitive performance, although slightly lower than the DeNN and RUN-XGBoost approaches. However, the results obtained with KNN and RF represent a significant improvement and highlight the applicability of these techniques in predicting the amount of ANFO in bench blasting (see Table 6).

Table 6. Comparison with existing models

Model	Previous Studies (R^2)	Current Study (R^2)	Reference
ANN-MLP	-	0.70	-
RF	-	0.78	-
SVR	-	0.55	-
XGBoost	0.98	0.72	[3]
KNN	0.87	0.84	[3]
RB	-	0.65	-
Direction-Encoded Neural Network (DeNN)	0.97	-	[47]
RUN-XGBoost	0.96	-	[48]

4. CONCLUSIONS

This research successfully employed various machine learning models, including ANN-MLP, RF, SVR, XGBoost, KNN, and BR, to predict the required ANFO quantity for bench blasting in an open-pit mine. The results demonstrated that KNN and RF models are particularly effective for ANFO quantity prediction. While the XGBoost, ANN-MLP, and BR models also performed well, there remains potential for further improvement through hyperparameter optimization.

Among the models tested, all six regression-based models exhibited satisfactory accuracy levels, with R^2 values above 0.55. The KNN model outperformed the others, achieving the highest accuracy with an R^2 of 0.84, an RMSE of 2.37, and an MAE of 1.35. The RF model followed closely, with an R^2 of 0.78, an RMSE of 2.77, and an MAE of 1.02. The XGBoost model recorded an R^2 of 0.72, an RMSE of 3.17, and an MAE of 1.35. Both ANN-MLP and BR models obtained very similar results, with R^2 values of 0.70 and 0.65, RMSE of 3.25 and 3.52, and MAE of 1.85 and 1.33, respectively. The SVR model, while still viable, had the lowest prediction accuracy, with an R^2 of 0.55, an RMSE of 4.01, and an MAE of 1.66.

It was identified that variables such as the height of the drill hole and the theoretical ANFO are critical determinants in the prediction of the quantity of ANFO required. The analysis established a direct relationship, showing that an increase in the height of the drill hole leads to an increase in the ANFO required. According to the KNN model, which proved to be the most accurate in this study, the quantity of ANFO required for bench blasting is estimated to be 13,322.6 kg.

The application of accurate models such as KNN and RF to predict the amount of ANFO can lead to cost savings by reducing explosive waste and preventing over-fragmentation of the rock. Additionally, these models enhance operational efficiency by ensuring more uniform fragmentation, which optimizes cycle times and reduces machinery wear. In terms of safety, precise prediction minimizes the risks associated with improper blasting, protecting both workers and nearby infrastructure. These combined benefits enhance profitability and safety in mining operations.

For future improvements, it is recommended to integrate additional geological and environmental data, apply more complex models such as deep neural networks, and explore ensemble methods like stacking to improve prediction accuracy. Additionally, automated hyperparameter optimization could further refine model performance across different operational contexts.

REFERENCES

[1] Bhandari, S. (1997). Engineering Rock Blasting Operations, UK. https://miningandblasting.wordpress.com/wp-content/uploads/2009/09/engineering-rock-blasting-operations_bhandari.pdf.

[2] Adhikari, G. (2000). Empirical methods for the calculation of the specific charge for surface blast design. Fragblast, 4: 19-33. https://doi.org/10.1080/13855140009408061

[3] Mishra, R., Kumar, A., Singh, B. (2023). High-speed motion analysis-based machine learning models for prediction and simulation of Flyrock in surface mines. Applied Sciences, 13(17): 9906.

https://doi.org/10.3390/app13179906
[4] Biessikirski, A., Dworzak, M., Twardosz, M. (2023). Composition of fumes and its influence on the general toxicity and applicability of mining explosives. Mining, 3(4): 605-617. https://doi.org/10.3390/mining3040033
[5] Beveridge, A. (1998). Forensic Investigation of Explosions. Taylor & Francis. https://www.routledge.com/Forensic-Investigation-of-Explosions/Beveridge/p/book/9780367778200.
[6] Hernandes, V.V., Franco, M.F., Santos, J.M., Melendez-Perez, J.J., de Morais, D.R., Rocha, W.F.C., Borges, R., de Souza, W., Zacca, J.J., Lograd, L.P.L., Eberlin, M.N., Correa, D.N. (2015). Characterization of ANFO explosive by high accuracy ESI(±)–FTMS with forensic identification on real samples by EASI(−)–MS. Forensic Science International, 249: 156-164. http://doi.org/10.1016/j.forsciint.2015.01.006
[7] (S&P), S.A. (2024). Global Commodity Insights. Recuperado el 20 de Marzo de 2024, https://www.spglobal.com/commodityinsights/en/ci/products/explosives-and-blasting-chemical-economics-handbook.html.
[8] Biegańska, J., Barański, K. (2022). Thermodynamic analysis of the possibility of using biomass as a component of high-energy materials. Energies, 15(15): 5624. https://doi.org/10.3390/en15155624
[9] Konya, C., Walter, E. (1991). Rock Blasting and Overbreak Control. Montville: National Highway Institute. https://miningandblasting.wordpress.com/wp-content/uploads/2009/09/rock-blasting-overbreak-control_konya.pdf.
[10] Jimeno, C., Jimeno, E., Carcedo, F. (1995). Drilling and Blasting of Rocks. CRC Press. https://www.crcpress.com/Drilling-and-Blasting-of-Rocks/Jimeno-Jimeno/p/book/9789054101994.
[11] Persson, P., Holmberg, R., Lee, J. (1993). Rock Blasting and Explosives Engineering. Washington D.C.: CRC Press. https://books.google.com.pe/books?id=sdLO5HESJwgC&printsec=frontcover#v=onepage&q&f=false.
[12] Widodo, S., Anwar, H., Syafitri, A. (2019). Comparative analysis of ANFO and emulsion application on overbreak and underbreak at blasting development activity in underground Deep Mill Level Zone (DMLZ) PT Freeport Indonesia. IOP Conference Series: Earth and Environmental Science, 279(1): 012001. https://doi.org/10.1088/1755-1315/279/1/012001
[13] Chen, Y., Wang, P., Chen, J., Zhou, M., Yang, H., Li, J. (2022). Calculation of blast hole charge amount based on three-dimensional solid model of blasting rock mass. Scientific Reports, 12(1): 541. https://doi.org/10.1038%2Fs41598-021-04615-8
[14] Duranović, M., Đokić, N., Lapčević, V., Torbica, S., Petrović, M., Savić, L. (2018). Optimization of ring blasting in sublevel stoping gold mine. Podzemni Radovi, (33): 61-68. http://doi.org/10.5937/PodRad1833061D
[15] Adhikari, G., Theresraj, A., Balachander, R., Gupta, R. (2001). Controlled blasting for removal of concrete plugs in draft tube tunnels at sardar Sarovar project. Fragblast, 5(4): 221-234. https://doi.org/10.1080/13855140009408061
[16] Bui, D., Bui, X., Drebenstedt, C., Nguyen, H. (2020). Prediction of blast-induced ground vibration in an open-pit mine by a novel hybrid model based on clustering and artificial neural network. Natural Resources Research, 29(2): 691-709. https://doi.org/10.1007/s11053-019-09470-z
[17] Bakhtavar, E., Sadiq, R., Hewage, K. (2021). Optimization of blasting-associated costs in surface mines using risk-based probabilistic integer programming and firefly algorithm. Natural Resources Research, 30(6): 4789-4806. https://doi.org/10.1007/s11053-021-09935-0
[18] Bayat, P., Monjezi, M., Mehrdanesh, A., Khandelwal, M. (2022). Blasting pattern optimization using gene expression programming and grasshopper optimization algorithm to minimise blast-induced ground vibrations. Engineering with Computers, 38(4): 3341-3350. https://doi.org/10.1007/s00366-021-01336-4
[19] Bayat, P., Monjezi, M., Rezakhah, M., Armaghani, D. (2020). Artificial neural network and firefly algorithm for estimation and minimization of ground vibration induced by blasting in a mine. Natural Resources Research, 29(6): 4121-4132. https://doi.org/10.1007/s11053-020-09697-1.
[20] Munagala, V., Thudumu, S., Logothetis, I., Bhandari, S., Vasa, R., Mouzakis, K. (2024). A comprehensive survey on machine learning applications for drilling and blasting in surface mining. Machine Learning with Applications, 15: 100517. https://doi.org/10.1016/j.mlwa.2023.100517
[21] Abd Elwahab, A., Topal, E. Jang, H. (2023). Review of machine learning application in mine blasting. Arabian Journal of Geosciences, 16(2): 133. https://doi.org/10.1007/s12517-023-11237-z
[22] Amoako, R., Jha, A., Zhong, S. (2022). Rock fragmentation prediction using an artificial neural network and support vector regression hybrid approach. Mining, 2(2): 233-247. https://doi.org/10.3390/mining2020013
[23] Rosales, J., Sáenz, R., Rojas, U., Castillo, J. (2020). Design of a predictive model of rock breakage by blasting using artificial neural networks. Symmetry, 12(9): 1405. https://doi.org/10.3390/sym12091405
[24] Ohadi, B., Sun, X., Esmaeili, K., Consens, M.P. (2019). Predicting blast-induced outcomes using random forest models of multi-year blasting data from an open pit mine. Bulletin of Engineering Geology and the Environment, 79(6): 329-343. http://doi.org/10.1007/s10064-019-01566-3
[25] Armaghani, D., Koopialipoor, M., Bahri, M., Hasanipaná, M., Tahir, M. (2020). A SVR-GWO technique to minimize flyrock distance resulting from blasting. Bulletin of Engineering Geology and the Environment, 79: 4369-4385. https://doi.org/10.1007/s10064-020-01834-7
[26] Nguyen, H., Bui, X., Bui, H., Trong, D. (2019). Developing an XGBoost model to predict blast-induced peak particle velocity in an open-pit mine: A case study. Acta Geophysica, 67: 477-490. https://doi.org/10.1007/s11600-019-00268-4
[27] Bui, X.N., Jaroonpattanapong, P., Nguyen, H., Tran, Q.H., Long, N.Q. (2019). A novel hybrid model for predicting blast-induced ground vibration based on K-Nearest Neighbors and particle swarm optimization. Scientific Reports, 9(1): 13971. https://doi.org/10.1038/s41598-019-50262-5
[28] Sun, P., Lu, W., Hu, H., Zhang, Y., Chen, M., Yan, P.

(2021). A Bayesian approach to predict blast-induced damage of high rock slope using vibration and sonic data. Sensors, 21(7): 2473. https://doi.org/10.3390%2Fs21072473

[29] Sola, J., Sevilla, J. (1997). Importance of input data normalization for the application of neural networks to complex industrial problems. IEEE Transactions on Nuclear Science, 44(3): 1464-1468. http://doi.org/10.1109/23.589532

[30] Larsen, R.J., Marx, M.L. (2005). An introduction to Mathematical Statistics. Hoboken, NJ: Prentice Hall. https://math.emory.edu/~lchen41/teaching/2020_Spring/Larsen-5E.pdf.

[31] Singh, D., Singh, B. (2020). Investigating the impact of data normalization on classification performance. Applied Soft Computing, 97: 105524. http://doi.org/10.1016/j.asoc.2019.105524

[32] Irie, Miyake. (1988). Capabilities of three-layered perceptrons. In IEEE 1988 International Conference on Neural Networks, San Diego, USA, pp. 641-648. https://doi.org/10.1109/ICNN.1988.23901

[33] Criminisi, A., Shotton, J., Konukoglu, E. (2012). Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. Computer Graphics and Vision, 7(2-3): 81-227. https://doi.org/10.1561/0600000035

[34] Trehan, S., Carlberg, K.T., Durlofsky, L.J. (2017). Error modeling for surrogates of dynamical systems using machine learning. International Journal for Numerical Methods in Engineering, 112(12): 1801-1827. https://doi.org/10.1002/nme.5583

[35] Fernández-Delgado, M., Cernadas, E., Barro, S., Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? The Journal of Machine Learning Research, 15(1): 3133-3181. https://jmlr.org/papers/v15/delgado14a.html.

[36] Ho, T.K. (1998). The random subspace method for constructing decision forests. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(8): 832-844. https://doi.org/10.1109/34.709601

[37] Breiman, L. (1996). Bagging predictors. Machine Learning, 24: 123-140. https://doi.org/10.1007/BF00058655

[38] Patel, A.K., Chatterjee, S., Gorai, A.K. (2019). Development of a machine vision system using the support vector machine regression (SVR) algorithm for the online prediction of iron ore grades. Earth Science Informatics, 12: 197-210. https://doi.org/10.1007/s12145-018-0370-6

[39] Rivas-Perea, P., Cota-Ruiz, J., Chaparro, D., Venzor, J., Carreón, A., Rosiles, J. (2013). Support vector machines for regression: A succinct review of large-scale and linear programming formulations. International Journal of Intelligence Science, 3: 5-14. https://doi.org/10.4236/ijis.2013.31002

[40] Chen, T., Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, California, San Francisco, USA, pp. 785-794. https://doi.org/10.48550/arXiv.1603.02754

[41] Yu, K., Ji, L., Zhang, X. (2002). Kernel nearest-neighbor algorithm. Neural Processing Letters, 15: 147-156. https://doi.org/10.1023/A:1015244902967

[42] Bárcena, M., Garín, A., Martín, A., Tusell, F., Unzueta, A. (2017). Un simulador para asistir en la ensenanza del teorema de Bayes. In In-Red 2017. III Congreso Nacional de innovación educativa y de docencia en red. Editorial Universitat Politècnica de València, pp. 15-23. https://doi.org/10.4995/INRED2017.2017.6725

[43] DataBitAI. (2023). Métricas de Evaluación en Machine Learning. https://databitai.com/machine-learning/metricas-de-evaluacion-en-machine-learning/#:~:text=Las%20m%C3%A9tricas%20de%20evaluaci%C3%B3n%20son,objetivos%20del%20problema%20a%20resolver.

[44] Prasad, K., Gorai, A., Goyal, P. (2016). Development of ANFIS models for air quality forecasting and input optimization for reducing the computational cost and time. Atmospheric Environment, 128: 246-262. https://doi.org/10.1016/j.atmosenv.2016.01.007

[45] Mahmoodzadeh, A., Reza, H., Mohammadi, M., Hashim, H., Khishe, M., Rashidi, S., Hama, H. (2022). Prediction of Mode-I rock fracture toughness using support vector regression with metaheuristic optimization algorithms. Engineering Fracture Mechanics, 264: 108334. https://doi.org/10.1016/j.engfracmech.2022.108334

[46] Sun, M., Yang, J., Yang, C., Wang, W., Wang, X., Li, H. (2024). Research on prediction of PPV in open-pit mine used RUN-XGBoost model. Heliyon, 10(7): e28246. https://doi.org/10.1016/j.heliyon.2024.e28246

[47] Dennis, A.A., Rigby, S.E. (2024). The Direction-encoded Neural Network: A machine learning approach to rapidly predict blast loading in obstructed environments. International Journal of Protective Structures, 15(3): 455-483. https://doi.org/10.1177/20414196231177364

[48] Majid, G., Nematollah, A., Saeid, R., Hamed, S. (2016). Prediction of blast boulders in open pit mines via multiple regression and artificial neural networks. International Journal of Mining Science and Technology, 26(2): 183-186. https://doi.org/10.1016/j.ijmst.2015.12.00