Machine Learning Algorithms for Data Enrichment: A Promising Solution for Enhancing Accuracy in Predicting Blast-Induced Ground Vibration in Open-Pit Mines

Hoang NGUYEN[1, 2)], Xuan-Nam BUI[1, 2)], Carsten DREBENSTEDT[3)]*

[1)] Surface Mining Department, Mining Faculty, Hanoi University of Mining and Geology, Hanoi 100000, Vietnam
[2)] Innovations for Sustainable and Responsible Mining (ISRM) Research Group, Hanoi University of Mining and Geology, Hanoi 100000, Vietnam
[3)] TU Bergakademie Freiberg, 09596 Freiberg, Sachen, Germany
* Corresponding author: nguyenhoang@humg.edu.vn (H. N.)

http://doi.org/10.29227/IM-2023-02-15

Submission date: 16-08-2023 | Review date: 22-09-2023

Abstract

The issue of blast-induced ground vibration poses a significant environmental challenge in open-pit mines, necessitating precise prediction and control measures. While artificial intelligence and machine learning models hold promise in addressing this concern, their accuracy remains a notable issue due to constrained input variables, dataset size, and potential environmental impact. To mitigate these challenges, data enrichment emerges as a potential solution to enhance the efficacy of machine learning models, not only in blast-induced ground vibration prediction but also across various domains within the mining industry. This study explores the viability of utilizing machine learning for data enrichment, with the objective of generating an augmented dataset that offers enhanced insights based on existing data points for the prediction of blast-induced ground vibration. Leveraging the support vector machine (SVM), we uncover intrinsic relationships among input variables and subsequently integrate them as supplementary inputs. The enriched dataset is then harnessed to construct multiple machine learning models, including k-nearest neighbors (KNN), classification and regression trees (CART), and random forest (RF), all designed to predict blast-induced ground vibration. Comparative analysis between the enriched models and their original counterparts, established on the initial dataset, provides a foundation for extracting insights into optimizing the performance of machine learning models not only in the context of predicting blast-induced ground vibration but also in addressing broader challenges within the mining industry.

Keywords: *blast-induced ground vibration, data enrichment, sustainable and responsible mining, machine learning, open-pit mining, performance improvement*

1. Introduction

Surface mining stands as one of the prevailing techniques for the exploitation of minerals, fossil fuels, and metals, characterized by its high degree of mechanization and productivity. Among the array of rock fragmentation methods employed in open-pit mining operations, drilling and blasting emerge as the most prevalent approach for fracturing rocks prior to subsequent unit operations like loading and hauling. The advantages of blasting are well-documented and undeniable; however, its detrimental repercussions, including blast-induced ground vibration (measured by peak particle velicity – PPV), flyrock, airblast, and air pollution [1-4], cannot be disregarded. Among these consequences, PPV is a particularly perilous phenomenon that exerts a profound impact on adjacent areas, notably open-pit mines situated in proximity to residential zones. Although efforts have been invested in assessing such hazards and proposing probabilistic risk-based models to manage these challenges, ensuring the safety of blasting operations [5], the complexity of blasting remains evident, encompassing a spectrum of potential accident risk scenarios [6]. Indeed, numerous structures have borne the brunt of PPV-induced cracks, and several slopes and benches have experienced subsidence or instability owing to the elevated magnitude of PPV within open-pit mines [7, 8]. Consequently, the accurate prediction of PPV intensity becomes a critical imperative, serving not only the preservation of neighboring structures but also the operational efficiency of open-pit mining ventures.

To achieve this goal, numerous researchers have put forth diverse predictive models aimed at estimating PPV. These models can be categorized into two primary groups: empirical models [8-11] and artificial intelligence (AI)-based models [12-19]. While empirical models have been endorsed despite their inherent accuracy limitations, AI-based models have garnered recognition for their exceptional performance. Presently, an increasing number of novel AI-based models have emerged, offering promising outcomes for PPV prediction, as well as other challenges not only within the realm of blasting but also across the broader mining industry [7, 20-34].

However, the majority of existing research has predominantly concentrated on enhancing predictive models using various techniques applied to raw datasets, or employing basic data analysis methods such as feature selection and outlier removal. In contrast, datasets containing more comprehensive and detailed information possess the potential to offer invaluable insights to predictive models. Such datasets can significantly aid predictive models in elucidating the intricate relationships between dependent and independent variables. Strikingly, these approaches appear to have been underexplored in the context of predicting the adverse effects of blast-

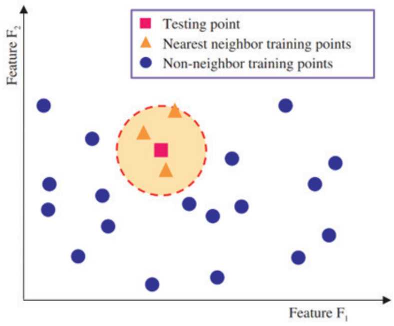


Fig. 1. Illustration of KNN algorithm for two-dimensional feature space [44]

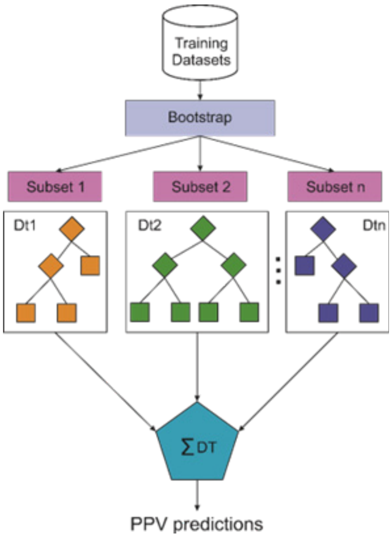


Fig. 2. Workflow of RF in predicting blast-induced PPV [52]

Fig. 3. Location and a view of the Coc Sau open-pit coal mine

ing in open-pit mines, including the prediction of PPV.

Hence, within the scope of this investigation, we introduce a novel approach focused on enhancing the PPV dataset as a precursor to the development of predictive models utilizing machine learning algorithms. Specifically, the support vector machine (SVM) algorithm was harnessed to discern the intricate relationships within the original PPV dataset's input variables. The resultant findings were then amalgamated with the initial dataset, culminating in the creation of an augmented dataset—referred to as the enriched dataset—containing a more expansive set of input variables. Subsequently, we crafted three distinct machine learning models: classification and regression trees (CART), k-nearest neighbors (KNN), and random forest (RF). These models were constructed employing both the original dataset and the enriched dataset, facilitating a comprehensive comparative analysis for PPV prediction within open-pit mining contexts. It is noteworthy that the enrichment technique proposed in this study is distinct from ensemble modeling approaches such as bagging, boosting, or stacking techniques [35-37].

2. Methodology

2.1. Data enrichment

Data enrichment refers to the process of enhancing or expanding the existing information or data sets by adding additional relevant details or attributes. It involves augmenting raw data with various types of supplemental information to make it more valuable and useful for analysis, soft computing models, decision-making, and other business purposes. The objective of data enrichment is to provide a more comprehensive and accurate understanding of the data by filling in missing gaps, correcting errors, or adding context. By enhancing the quality and depth of data, organizations can gain deeper insights, improve customer understanding, and make more informed decisions.

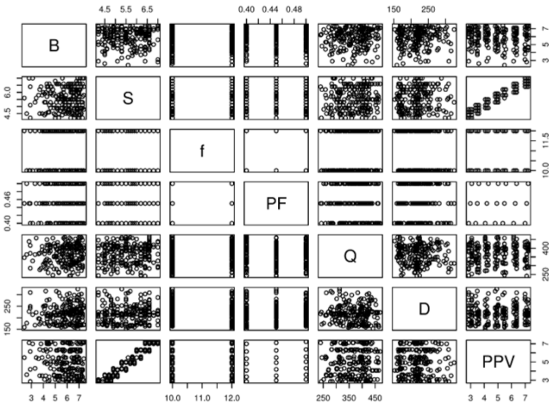


Fig. 4. Visualization of the original dataset used in this study

Tab. 1. Statistical parameters of the original dataset

B	S	f	PF	Q	D	PPV
Min. :2.400	Min. :4.200	Min. :10.00	Min. :0.400	Min. :240.0	Min. :152.4	Min. :2.86
1st Qu.:5.225	1st Qu.:4.900	1st Qu.:10.00	1st Qu.:0.400	1st Qu.:319.5	1st Qu.:189.3	1st Qu.:3.60
Median :6.200	Median :5.600	Median :12.00	Median :0.450	Median :377.5	Median :218.2	Median :4.92
Mean :5.893	Mean :5.598	Mean :11.01	Mean :0.452	Mean :369.0	Mean :219.2	Mean :4.97
3rd Qu.:6.800	3rd Qu.:6.400	3rd Qu.:12.00	3rd Qu.:0.500	3rd Qu.:416.0	3rd Qu.:241.8	3rd Qu.:6.27
Max. :7.400	Max. :7.000	Max. :12.00	Max. :0.500	Max. :469.0	Max. :315.1	Max. :7.22

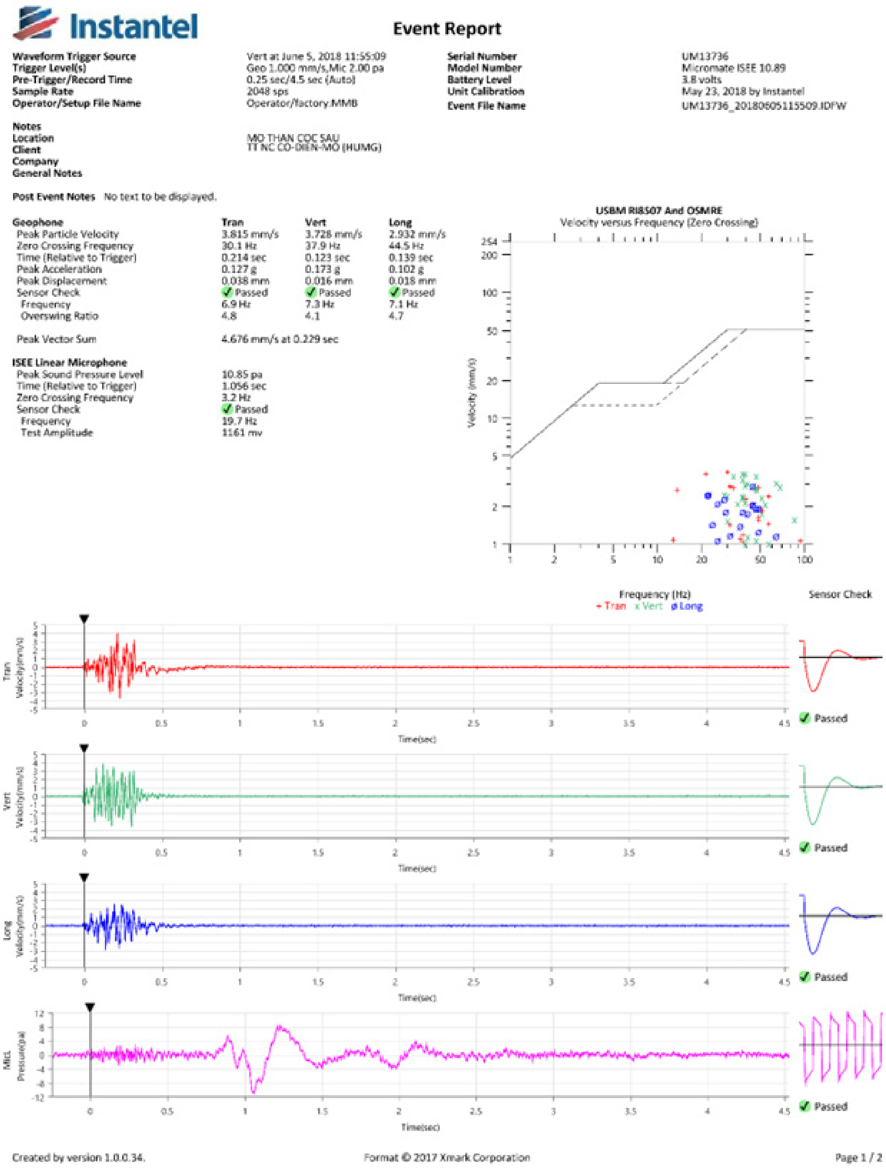


Fig. 5. PPV resulting from blasting operation at the Coc Sau open-pit coal mine

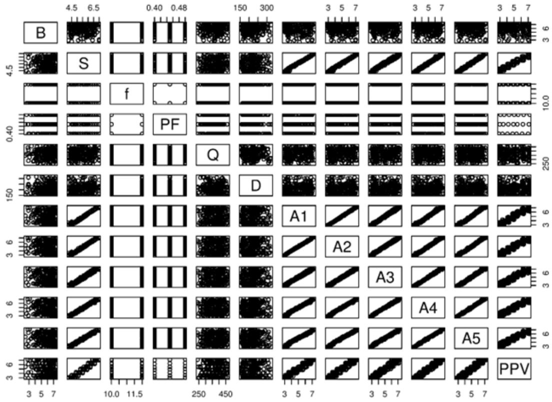


Fig. 6. Visualization of the enriched dataset used in this study

Tab. 1. Statistical parameters of the original dataset

B	S	f	PF	Q	D
Min. :2.400	Min. :4.200	Min. :10.00	Min. :0.4000	Min. :240.0	Min. :152.4
1st Qu.:5.175	1st Qu.:4.900	1st Qu.:10.00	1st Qu.:0.4000	1st Qu.:318.5	1st Qu.:184.4
Median :6.150	Median :5.500	Median :12.00	Median :0.4500	Median :370.5	Median :217.5
Mean :5.854	Mean :5.582	Mean :11.03	Mean :0.4528	Mean :365.9	Mean :218.1
3rd Qu.:6.800	3rd Qu.:6.250	3rd Qu.:12.00	3rd Qu.:0.5000	3rd Qu.:410.2	3rd Qu.:239.8
Max. :7.400	Max. :7.000	Max. :12.00	Max. :0.5000	Max. :469.0	Max. :324.7
A1	**A2**	**A3**	**A4**	**A5**	**PPV**
Min. :2.850	Min. :2.778	Min. :2.678	Min. :2.640	Min. :2.648	Min. :2.860
1st Qu.:3.600	1st Qu.:3.690	1st Qu.:3.694	1st Qu.:3.754	1st Qu.:3.766	1st Qu.:3.600
Median :4.842	Median :4.870	Median :4.940	Median :4.900	Median :4.881	Median :4.920
Mean :4.959	Mean :4.959	Mean :4.960	Mean :4.934	Mean :4.930	Mean :4.961
3rd Qu.:6.197	3rd Qu.:6.138	3rd Qu.:6.147	3rd Qu.:6.100	3rd Qu.:6.036	3rd Qu.:6.270
Max. :7.328	Max. :7.253	Max. :7.367	Max. :7.283	Max. :7.281	Max. :7.220

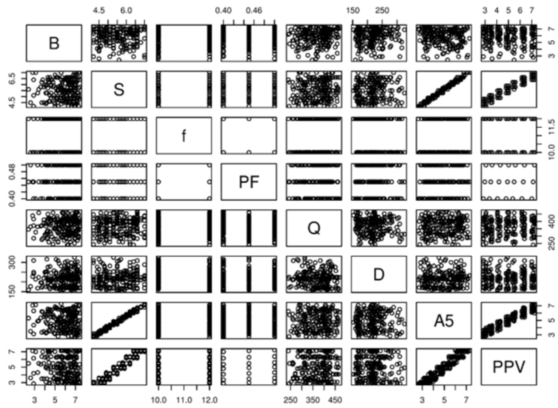


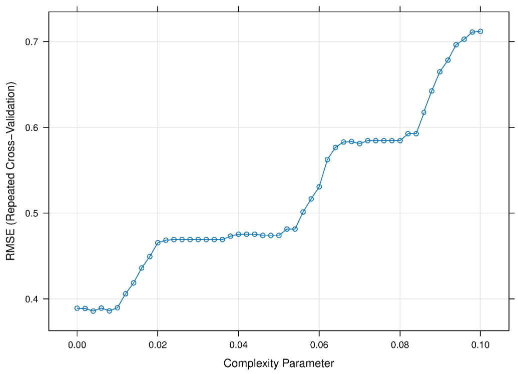
Fig. 7. Visualization of the final enriched dataset used (after analyzing) in this study

Data enrichment plays a crucial role in various domains, especially in engineering problems, such as mining, civil engineering, geotechnical engineering, material engineering, mechanics, to name a few. By leveraging the power of enriched data, researchers can explore the insights of the datasets, optimize operations, or enhancing problems they encountered. There are several techniques that can be used to enrich data for machine learning, including feature engineering, data augmentation, imputation, oversampling and undersampling, feature selection, and external data integration. Of those, data augmentation is a technique used to artificially increase the size and diversity of a dataset by applying various transformations to existing data samples. It is commonly used to enrich the dataset for machine learning tasks. While data augmentation is primarily applied to address challenges in computer vision tasks, such as image classification, object detection, or segmentation, it can also be adapted for other types of data, including regression and time series data.

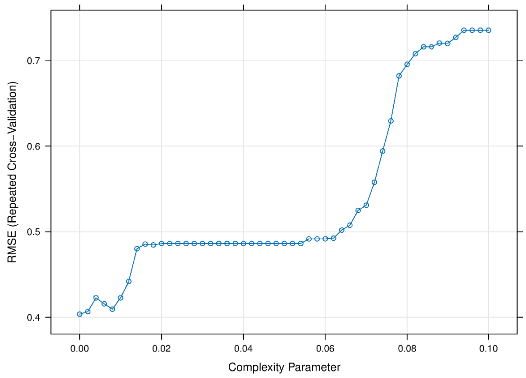
In this study, the SVM machine learning model was used for data enrichment purposes. As a matter of fact, other machine learning algorithms can also do the same; however, in this study, we selected the SVM algorithm as its popularity and simple. Although the SVM is well-known as a black-box model, however, it can explain the relationships between independent variables through a function.

Assume that we have a dataset contains eight input variables, named as $X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8$, and the response variable is Y. The relationships between the input variables can be expressed through different functions that combining the input variables together, as follow:

$$Y_1 = f_{SVM}(X_1, X_2) \quad (1)$$
$$Y_2 = f_{SVM}(X_1, X_2, X_3) \quad (2)$$
$$Y_3 = f_{SVM}(X_1, X_2, X_3, X_4) \quad (3)$$
$$Y_4 = f_{SVM}(X_1, X_2, X_3, X_4, X_5) \quad (4)$$
$$Y_5 = f_{SVM}(X_1, X_2, X_3, X_4, X_5, X_6) \quad (5)$$
$$Y_6 = f_{SVM}(X_1, X_2, X_3, X_4, X_5, X_6, X_7) \quad (6)$$

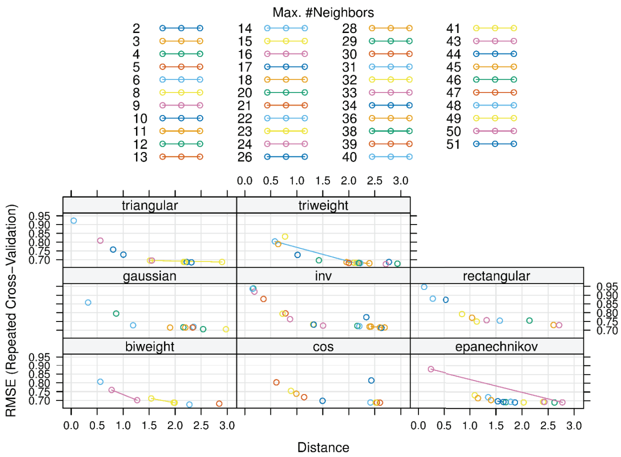


(a) On the original dataset

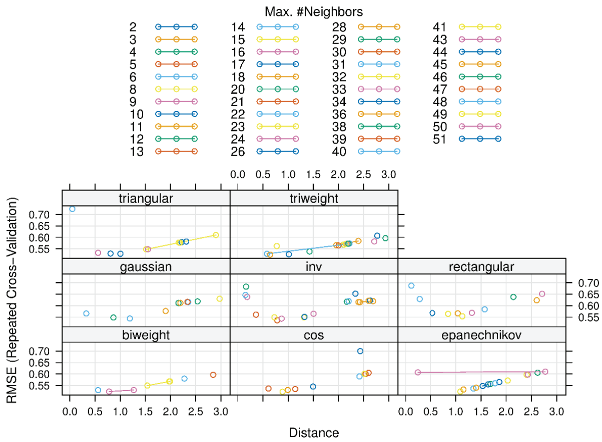


(b) On the enriched dataset

Fig. 8. Performance of the CART model for predicting PPV on the original dataset and enriched dataset

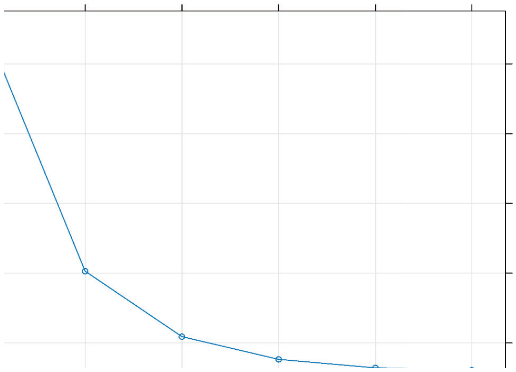


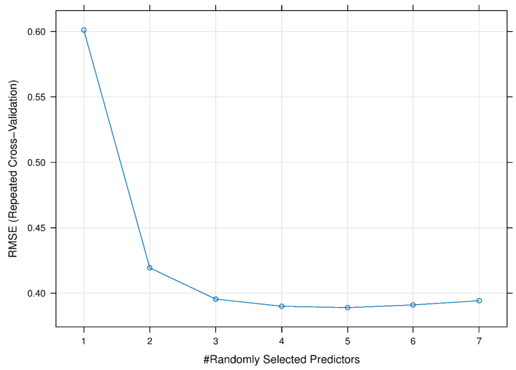
(a)O nt he original dataset



(b)O nt he enriched dataset

Fig. 9. Performance of the KNN model for predicting PPV on the original dataset and enriched dataset


(a) On the original dataset



(b) On the enriched dataset

Fig. 10. Performance of the RF model for predicting PPV on the original dataset and enriched dataset

Tab. 3. Performance of the PPV predictive models based on the original dataset and enriched dataset

Model	Training dataset			Testing dataset		
	RMSE	R^2	MAE	RMSE	R^2	MAE
Original dataset						
CART	0.385	0.934	0.351	0.405	0.911	0.360
KNN	0.674	0.811	0.571	0.640	0.798	0.544
RF	0.361	0.944	0.312	0.351	0.935	0.293
Enriched dataset						
CART	0.404	0.929	0.351	0.392	0.921	0.332
KNN	0.522	0.878	0.451	0.438	0.903	0.354
RF	0.389	0.932	0.348	0.343	0.938	0.309

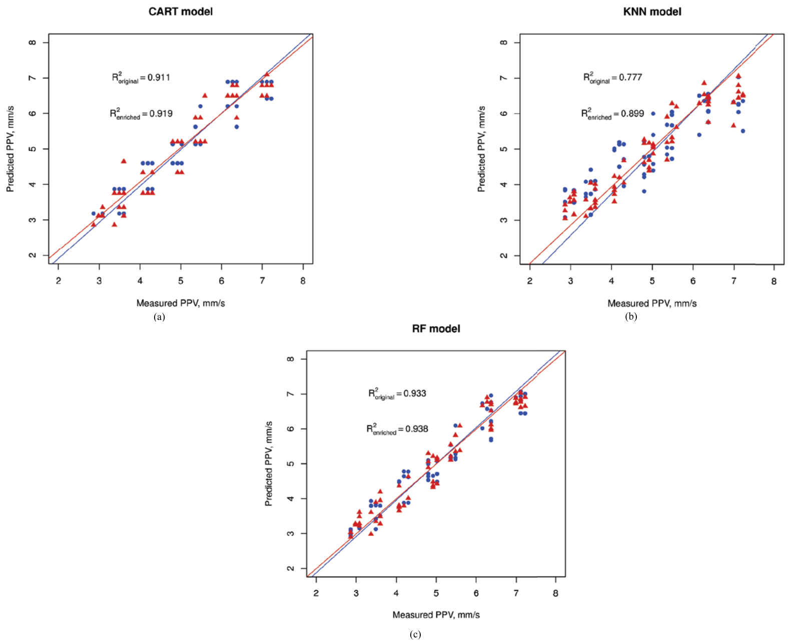


Fig. 11. Correlation between the actual and predicted PPV values on both original dataset and enriched dataset: (a) CART model; (b) KNN model; (c) RF model

$$Y_7 = f_{SVM}(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8) \tag{7}$$

Through the seven functions above, the relationships between input variables are explained. Further details of the dataset has been explored through the SVM functions. Finally, they are added to the original dataset as the additional input variables to provide more detail of the dataset's information. In other words, the SVM functions based on different combination of input variables artificially increased the size and diversity of the assumed dataset. The principle of the SVM model is brief described as below.

The SVM, a machine learning algorithm introduced by Cortes and Vapnik [38], is designed to minimize structural risk, enabling better generalization from a limited set of samples. SVM has the ability to tackle classification and regression tasks. In the context of regression, it is referred to as Support Vector Regression (SVR), which constructs a forecasting model based on a subset of the training dataset [39].

To predict blast-induced PPV, SVR can be executed using one of the kernel functions listed below:

- Linear kernel:

$$K(x,y) = x \cdot y \tag{8}$$

- Polynomial kernel:

$$K(x,y) = [(x \cdot y)+1]^d \;\; ;d = (1,2,...) \tag{9}$$

- Radial primary kernel function:

$$K(x,y) = \exp[\frac{-\|x-y\|^2}{\sigma^2}] \tag{10}$$

- Two-layer neural kernel:

$$K(x,y) = \tanh[a(x \cdot y) - \delta] \tag{11}$$

2.2. Machine learning models for predicting PPV

2.2.1. k-nearest neighbors (KNN)

KNN, proposed by Altman [40], stands out as one of the simplest supervised-learning algorithms in AI. It falls under the category of lazy machine learning since it doesn't learn from training datasets. Instead, KNN makes predictions for new data points based on computations conducted using existing data. This instance-based or memory-based learning algorithm is versatile, supporting both classification and regression tasks.

For classification problems, KNN determines the output of a data point by looking at the nearest known data point (k = 1) or the weighted average of the closest neighbors' outputs. In regression problems, the output is calculated based on the relationship with the nearest data point, depending on the distance.

In essence, KNN predicts the output for a new data point solely based on the information from k data points in the closest training set (k-neighborhood), disregarding any interference from the surrounding data points. For more in-depth details about the KNN algorithm, refer to Song, Liang [41] and Chae, Lee [42].

Interestingly, KNN has been recommended as one of good solutions for predicting blast-induced PPV in open-pit mines [43]. Therefore, this study explores its application for this specific purpose. The next section delves into the process of determining the number of neighbors and setting up the KNN model. Figure 1 presents the mechanism of the KNN model.

2.2.2. Classification and regression trees (CART)

The CART algorithm, widely used in statistical communities, is a non-parametric decision tree algorithm used to predict dependent variables based on independent variables [45]. Inspired by the growth of trees in nature, the CART decision tree operates by segregating independent variables into homogeneous regions, characterized by roots, leaves, branches, and nodes [46, 47].

Breiman, Friedman [48] describe CART as an estimating method that doesn't rely on initial hypotheses about the relationship between dependent and independent variables. It efficiently identifies significant variables while discarding unimportant ones and demonstrates excellent handling of outliers, which can be detrimental to statistical models. The key features of the CART algorithm are as follows:

- Data extraction at a node is based on the value of a specific variable, applying predefined rules.
- It utilizes specific criteria to control the creation of complex trees.
- Pruning is employed to optimize the model's performance.
- The algorithm calculates and predicts the output for terminal nodes.

For this particular study, the CART algorithm was chosen as the benchmark regression algorithm to predict seismic vibration caused by blasting in fragmenting rock. The next sections provide detailed explanations of the CART model setup and PPV forecasts.

2.2.3. Random forest (RF)

RF, proposed by Breiman [49], is an ensemble machine learning algorithm belonging to the group of decision tree algorithms. It is versatile, capable of solving both classification and regression problems. The essence of RF lies in constructing multiple decision trees through bootstrap aggregation (bagging) [50, 51]. It combines the results from these trees to make a final decision. Each tree is trained with a random selection of variables and data samples from the initial training dataset.

For the prediction of blast-induced PPV, RF was applied as follows:

- The number of trees was chosen to ensure a rich forest.
- Bootstrap samples were drawn with replacement from the original PPV training dataset. The remaining values were used for validation and referred to as out-of-bag (OOB) data.
- A non-pruning regression was developed with modifications at each node for each bootstrap sample.
- At each bootstrap iteration, OOB data was used to predict PPV, and the results were averaged across all trees.
- Performance indices such as RMSE, R2, and MAE were used to evaluate the accuracy of predicted PPV values on OOB.

3. Data preparation and model development

3.1. Original datasets

In this study, a dataset consists of 216 blasting events was collected at the Coc Sau open-pit coal mine (Vietnam). This is the deepest open-pit coal mine in Vietnam (-300 m below sea level), as shown in Figure 3.

With the hardness of rock is in the range of 10 to 14, blasting is still the most effective method for fragmenting rock in this open-pit mine. Herein, the dataset with the parameters, such as burden (B), spacing (S), rock hardness (f), powder factor (PF), maximum explosive charge per delay (Q), and PPV monitoring distance (D), were collected and measured for predicting PPV at the Coc Sau open-pit coal mine. The details of the original dataset is shown in Table 1 and its visualization is shown in Figure 4.

Accordingly, B, S, f, PF, and Q were exported from the blast patterns, and D was measured by a GPS receiver from the blast sites to the geo-phone blasting. PPV was measured by the Micromate device (Instantel – Canada). Figure 5 shows a result of PPV that was measured by Micromate at the Coc Sau open-pit coal mine.

3.2. Data enrichment

As previously introduced, the SVM algorithm was employed to elucidate the correlations among the input variables, as outlined in Equations (1-7). Within this research, the original dataset encompassed 6 input variables. The SVM algorithm dissected these relationships through 5 distinct SVM models, namely: the SVM model involving the B and S variables; the SVM model involving the B, S, and F variables; the SVM model involving the B, S, f, and PF variables; the SVM model involving the B, S, f, PF, and q variables; and the SVM model involving the B, S, f, PF, q, and S variables. For each model, an additional novel variable was generated based on the SVM model's predictive outcomes. Consequently, this yielded 5 supplementary variables (A1, A2, A3, A4, A5), augmenting the original dataset to a more comprehensive state featuring 11 input variables. The particulars of this enriched dataset are presented in Table 2, with its visualization depicted in Figure 6.

The findings depicted in Figure 6 reveal noteworthy correlations between five supplementary variables. Consequently, the initial four supplementary variables were eliminated from the augmented dataset. The ultimate supplementary variable, however, was retained due to its capability to elucidate the interrelationships among the remaining variables. As a result, the operational dataset encompasses seven input variables. It's important to note that the final input variable was generated via the SVM algorithm, leveraging the inherent relationships among the original input variables. This enrichment process significantly enhanced the dataset's informational content compared to its original state. The resulting enriched dataset, subsequent to analysis, is presented visually in Figure 7.

4. Results and discussion

After successfully completing the enrichment process, the KNN, CART, and RF models were developed for PPV prediction using both the original and enriched datasets. Subsequently, the predictive outcomes were compared before and after the enrichment process to assess the effectiveness of the SVM algorithm in enhancing the PPV dataset.

In preparation for constructing the predictive models, both the initial and enriched datasets were split into two equal parts following a 70/30 ratio. This allocation designated 70% of each dataset for model training, while the remaining 30% was reserved to evaluate model performance in practical scenarios. To avert overfitting during model development, techniques such as 5-fold cross-validation with 3 repetitions, Box-Cox transformation, and center scaling were implemented.

It's important to emphasize that all models employed identical training and testing datasets for both training and validation phases. Within this study, the Root-mean-squared error (RMSE) function was adopted as the loss metric while training the CART, KNN, and RF models for PPV prediction.

For the CART model, the complexity parameter was used to control the accuracy of this model, and a grid search in the range of 0.002 to 0.1 was used to determine the optimal parrameter of the CART model on the original dataset and enriched dataset. The training results on the original dataset and enriched dataset are shown in Figure 8.

For the KNN model, three parameters, including the maximum number of neighbors (k), the distance beween the nearest neighbors, and kernel function were used to control the accuracy of the accuracy of the KNN model. A random search technique with 100 different KNN models based on 100 different set of parameters were emplemented on both the original dataset and enriched dataset, as shown in Figure 9.

For the RF model, the number of trees in the forest and random selected predictors were used to control the accuracy of this model. According to the experience of previous researchers, a diversity of trees in the forest can improve the prediction performance of the RF model. Therefore, it was selected equal to 2000 trees in this study. In addition, a grid search in the range of 1 to 6 for the original dataset, and a range of 1 to 7 for the enricheed dataset, were applied during developing the RF model for predicting PPV in this study. The training results are shown in Figure 9.

Once the CART, KNN and RF models were developed on both original dataset and enriched dataset, the testing datasets of both original and enriched datasets were used to validate the performance of the developed models. Performance metrics, including RMSE, determination coefficient (R^2) and mean absolute error (MAE) were calculated according to equations (12-14) to evaluate the performance of the developed models, as shown in Table 3.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n_{blast}} \sum_{blast=1}^{n_{blast}} (PPV_i - \hat{P}PV_i)^2} \tag{12}$$

$$R^2 = 1 - \frac{\sum_{blast=1}^{n_{blast}} \left(PPV_i - \hat{P}PV_i\right)^2}{\sum_{blast=1}^{n_{blast}} \left(PPV_i - \overline{PPV}_i\right)^2} \tag{13}$$

$$\mathrm{MAE} = \frac{1}{n_{blast}} \sum_{blast=1}^{n_{blast}} \left| PPV_i - \hat{P}PV_i \right| \tag{14}$$

where n_{blast} is the number of blasting cases used in the dataset; PPV_i, PPV_i, and PPV_i stand for the measured PPV, predicted PPV, and mean of measured PPVs.

Across all models and datasets, the RMSE values reflect the average magnitude of prediction errors. Smaller RMSE values indicate better predictive accuracy. R^2 values provide an indication of how well the model's predictions fit the actual data. Higher R^2 values suggest a better fit. MAE measures the average absolute difference between predictions and actual values. Lower MAE values indicate better accuracy.

Based on the obtained results in Table 3 and comparing the original and enriched datasets, there are noticeable changes in the performance metrics. The RMSE values in the enriched dataset are indeed smaller, indicating better predictive accuracy compared to the original dataset. Of those, the RF model consistently performs well on both datasets in terms of RMSE, R^2, and MAE, suggesting its robustness in this context. Also, the KNN model demonstrates improvements in performance on the enriched dataset, particularly in terms of RMSE and R^2. And the CART model's performance remains relatively stable between the two datasets, and its performance on the enriched dataset is also slightly better than those of the original dataset. Figure 10 shows the correlation between the actual and predicted PPV values on both original dataset and enriched dataset.

As illustrated in Figure 10, the proximity between the projected outcomes and the actual PPV measurements is notably enhanced in the results obtained from the enriched dataset, in contrast to the projected PPV values originating from the original dataset. This implies that the predictive models exhibited greater convergence when applied to the enriched dataset compared to the original dataset. This observation strongly suggests that the enrichment process substantially bolstered the predictive models' efficacy in forecasting PPV within the scope of this study.

5. Conclusion

Blasting constitutes a pivotal component of surface mining technology, yet its ramifications, notably the considerable impact of blast-induced ground vibration (quantified as PPV in this study), wield a significant influence on the surrounding environment, warranting precise prediction and control. To address this issue, two viable approaches have been identified:

Enhancing Predictive Models: One strategy involves refining predictive models through the implementation of diverse optimization techniques or clustering methodologies.

Dataset Enrichment: Another avenue involves augmenting the dataset to furnish more intricate details, thereby enabling predictive models to forecast PPV with heightened accuracy.

In the context of this investigation, we have proposed a promising remedy to augment the precision of PPV predictive models (specifically CART, KNN, and RF models) employed in open-pit mining settings. This solution involves leveraging machine learning algorithms, notably the SVM algorithm. The outcomes garnered underscore the potential of machine learning algorithms in elucidating the interplay among input variables within the original dataset. The resultant insights can then be employed as supplementary variables to enhance the original dataset, thereby facilitating the improved performance of predictive models in PPV prognostication. Utilizing this data enrichment technique, novel AI-based models can enhance their accuracy to a greater extent compared to utilizing the original dataset alone.

Acknowledgments

Dr. Hoang Nguyen was funded by the Postdoctoral Scholarship Programme of Vingroup Innovation Foundation (VINIF), code VINIF.2022.STS.23. The authors would like to thank the organization committee of the 7th International Conference POL-VIET, Scientific and research cooperation between Vietnam and Poland.

Literatura – References

1. Abbaszadeh, S.A. and R. Asheghi, Optimized developed artificial neural network-based models to predict the blast-induced ground vibration. Innovative Infrastructure Solutions, 2018. 3: p. 1-10.
2. Ainalis, D., et al., Modelling the source of blasting for the numerical simulation of blast-induced ground vibrations: a review. Rock Mechanics and Rock Engineering, 2017. 50(1): p. 171-193.
3. Monjezi, M., et al., Predicting blast-induced ground vibration using various types of neural networks. Soil Dynamics and Earthquake Engineering, 2010. 30(11): p. 1233-1236.
4. Marks, N.A., et al., Airblast variability and fatality risks from a VBIED in a complex urban environment. Reliability Engineering & System Safety, 2021. 209: p. 107459.
5. Stewart, M.G. and M.D. Netherton, A probabilistic risk-acceptance model for assessing blast and fragmentation safety hazards. Reliability Engineering & System Safety, 2019. 191: p. 106492.
6. Gerassis, S., et al., Understanding complex blasting operations: A structural equation model combining Bayesian networks and latent class clustering. Reliability Engineering & System Safety, 2019. 188: p. 195-204.
7. Nguyen, H., et al., Predicting Blast-Induced Ground Vibration in Open-Pit Mines Using Different Nature-Inspired Optimization Algorithms and Deep Neural Network. Natural Resources Research, 2021. 30(6): p. 4695-4717.
8. Nicholls, H.R., C.F. Johnson, and W.I. Duvall, Blasting vibrations and their effects on structures. 1971: US Government Printers.
9. Murmu, S., P. Maheshwari, and H.K. Verma, Empirical and probabilistic analysis of blast-induced ground vibrations. International Journal of Rock Mechanics and Mining Sciences, 2018. 103: p. 267-274.
10. Mesec, J., I. Kovač, and B. Soldo, Estimation of particle velocity based on blast event measurements at different rock units. Soil Dynamics and Earthquake Engineering, 2010. 30(10): p. 1004-1009.
11. Ak, H., et al., Evaluation of ground vibration effect of blasting operations in a magnesite mine. Soil Dynamics and Earthquake Engineering, 2009. 29(4): p. 669-676.
12. Monjezi, M., M. Hasanipanah, and M. Khandelwal, Evaluation and prediction of blast-induced ground vibration at Shur River Dam, Iran, by artificial neural network. Neural Computing and Applications, 2013. 22(7): p. 1637-1643.
13. Saadat, M., M. Khandelwal, and M. Monjezi, An ANN-based approach to predict blast-induced ground vibration of Gol-E-Gohar iron ore mine, Iran. Journal of Rock Mechanics and Geotechnical Engineering, 2014. 6(1): p. 67-76.
14. Hasanipanah, M., et al., Feasibility of indirect determination of blast induced ground vibration based on support vector machine. Measurement, 2015. 75: p. 289-297.
15. Amiri, M., et al., A new combination of artificial neural network and K-nearest neighbors models to predict blast-induced ground vibration and air-overpressure. Engineering with Computers, 2016. 32(4): p. 631-644.
16. Azimi, Y., S.H. Khoshrou, and M. Osanloo, Prediction of blast induced ground vibration (BIGV) of quarry mining using hybrid genetic algorithm optimized artificial neural network. Measurement, 2019. 147: p. 106874.
17. Shang, Y., et al., A Novel Artificial Intelligence Approach to Predict Blast-Induced Ground Vibration in Open-Pit Mines Based on the Firefly Algorithm and Artificial Neural Network. Natural Resources Research, 2020. 29(2): p. 723-737.
18. Zhang, X., et al., Novel Soft Computing Model for Predicting Blast-Induced Ground Vibration in Open-Pit Mines Based on Particle Swarm Optimization and XGBoost. Natural Resources Research, 2020. 29(2): p. 711-721.
19. Qiu, Y., et al., Performance evaluation of hybrid WOA-XGBoost, GWO-XGBoost and BO-XGBoost models to predict blast-induced ground vibration. Engineering with Computers, 2021.
20. Bui, X.-N., H.-B. Bui, and H. Nguyen. A Review of Artificial Intelligence Applications in Mining and Geological Engineering. in Proceedings of the International Conference on Innovations for Sustainable and Responsible Mining. 2021. Springer.
21. Jung, D. and Y. Choi, Systematic review of machine learning applications in mining: Exploration, exploitation, and reclamation. Minerals, 2021. 11(2): p. 148.
22. Kim, M., L. Ismail, and S. Kwon, Review of the Application of Artificial Intelligence in Blasting Area. Explosives and Blasting, 2021. 39(3): p. 44-64.
23. Yong, W., et al., Analysis and prediction of diaphragm wall deflection induced by deep braced excavations using finite element method and artificial neural network optimized by metaheuristic algorithms. Reliability Engineering & System Safety, 2022. 221: p. 108335.
24. Liu, D., S. Wang, and X. Cui, An artificial neural network supported Wiener process based reliability estimation method considering individual difference and measurement error. Reliability Engineering & System Safety, 2022. 218: p. 108162.
25. Yang, Z., P. Baraldi, and E. Zio, A method for fault detection in multi-component systems based on sparse autoencoder-based deep neural networks. Reliability Engineering & System Safety, 2022. 220: p. 108278.

26. Seo, S.-K., et al., Deep neural network-based optimization framework for safety evacuation route during toxic gas leak incidents. Reliability Engineering & System Safety, 2022. 218: p. 108102.

27. Zhou, J., et al., Developing a hybrid model of Jaya algorithm-based extreme gradient boosting machine to estimate blast-induced ground vibrations. International Journal of Rock Mechanics and Mining Sciences, 2021. 145: p. 104856.

28. Lawal, A.I., et al., Blast-induced ground vibration prediction in granite quarries: An application of gene expression programming, ANFIS, and sine cosine algorithm optimized ANN. International Journal of Mining Science and Technology, 2021. 31(2): p. 265-277.

29. Zhu, W., H.N. Rad, and M. Hasanipanah, A chaos recurrent ANFIS optimized by PSO to predict ground vibration generated in rock blasting. Applied Soft Computing, 2021. 108: p. 107434.

30. Zhang, X., et al., Novel Extreme Learning Machine-Multi-Verse Optimization Model for Predicting Peak Particle Velocity Induced by Mine Blasting. Natural Resources Research, 2021. 30(6): p. 4735-4751.

31. Armaghani, D.J., et al., A novel approach for forecasting of ground vibrations resulting from blasting: modified particle swarm optimization coupled extreme learning machine. Engineering with computers, 2021. 37(4): p. 3221-3235.

32. Arthur, C.K., V.A. Temeng, and Y.Y. Ziggah, A Self-adaptive differential evolutionary extreme learning machine (SaDE-ELM): a novel approach to blast-induced ground vibration prediction. SN Applied Sciences, 2020. 2(11): p. 1845.

33. Ding, X., et al., Predicting the blast-induced vibration velocity using a bagged support vector regression optimized with firefly algorithm. Engineering with Computers, 2020: p. 1-12.

34. Nguyen, H., X.-N. Bui, and E. Topal, Reliability and availability artificial intelligence models for predicting blast-induced ground vibration intensity in open-pit mines to ensure the safety of the surroundings. Reliability Engineering & System Safety, 2023. 231: p. 109032.

35. Syarif, I., et al. Application of bagging, boosting and stacking to intrusion detection. in Machine Learning and Data Mining in Pattern Recognition: 8th International Conference, MLDM 2012, Berlin, Germany, July 13-20, 2012. Proceedings 8. 2012. Springer.

36. Graczyk, M., et al. Comparison of bagging, boosting and stacking ensembles applied to real estate appraisal. in Intelligent Information and Database Systems: Second International Conference, ACIIDS, Hue City, Vietnam, March 24-26, 2010. Proceedings, Part II 2. 2010. Springer.

37. Wolpert, D.H. and W.G. Macready, Combining stacking with bagging to improve a learning algorithm. Santa Fe Institute, Technical Report, 1996. 30.

38. Cortes, C. and V. Vapnik, Support vector machine. Machine learning, 1995. 20(3): p. 273-297.

39. Basak, D., S. Pal, and D.C. Patranabis, Support vector regression. Neural Information Processing-Letters and Reviews, 2007. 11(10): p. 203-224.

40. Altman, N.S., An introduction to kernel and nearest-neighbor nonparametric regression. The American Statistician, 1992. 46(3): p. 175-185.

41. Song, Y., et al., An efficient instance selection algorithm for k nearest neighbor regression. Neurocomputing, 2017. 251: p. 26-34.

42. Chae, D.-K., et al., On identifying k-nearest neighbors in neighborhood models for efficient and effective collaborative filtering. Neurocomputing, 2018. 278: p. 134-143.

43. Bui, X.-N., et al., A novel Hybrid Model for predicting Blast-induced Ground Vibration Based on k-nearest neighbors and particle Swarm optimization. Scientific reports, 2019. 9(1): p. 1-14.

44. Hu, C., et al., Data-driven method based on particle swarm optimization and k-nearest neighbor regression for estimating capacity of lithium-ion battery. Applied Energy, 2014. 129: p. 49-55.

45. Khandelwal, M., et al., Classification and regression tree technique in estimating peak particle velocity caused by blasting. Engineering with Computers, 2017. 33(1): p. 45-53.

46. Myles, A.J., et al., An introduction to decision tree modeling. Journal of Chemometrics, 2004. 18(6): p. 275-285.

47. Pradhan, B., A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. Computers & Geosciences, 2013. 51: p. 350-365.

48. Breiman, L., et al., Classification and decision trees. Wadsworth, Belmont, 1984. 378.

49. Breiman, L., Random forests. Machine learning, 2001. 45(1): p. 5-32.

50. Effron, B. and R.J. Tibshirani, An introduction to the bootstrap. Monographs on statistics and applied probability, 1993. 57: p. 436.

51. Breiman, L., Random forests. UC Berkeley TR567, 1999.

52. Nguyen, H., et al., Developing an XGBoost model to predict blast-induced peak particle velocity in an open-pit mine: a case study. Acta Geophysica, 2019. 67(2): p. 477-490.