Geotechnical parameters prediction from geophysical logging data using supervised learning methods

Gabriela Frainz
Vale S.A., Nova Lima, Brazil

Gustavo Gouveia
Vale S.A., Nova Lima, Brazil

ABSTRACT: Rotary percussion drilling, despite being more agile and economical, does not provide the recovery of intact material as in traditional methods. With the use of this type of drilling, obtaining geotechnical parameters from alternative methodologies, such as geophysical logging, becomes of fundamental importance. Currently, however, geophysical logging does not meet the need to describe cores obtained by conventional drilling, due to methodologies established for geotechnics are based on tactile-visual parameters. Therefore, an opportunity is identified in the application of machine learning algorithms that use different data sources to obtain geotechnical parameters. Three algorithms were applied to the data sources and the scores varied depending on the comparisons and the acceptance criteria. It was observed that when the adjustment sequence is made in a more flexible criteria of acceptance, high levels of accuracy are achieved, which makes this tool practicable in geotechnical projects.

Keywords: Machine Leaning, Geophysical, Geotechnical Logging, Mining, Supervised Learning.

1 INTRODUCTION

A competitive method that was designed to improve rock-breaking efficiency in hard formations (Xi et al., 2022), rotary percussion drilling has been widely recognized for its velocity, improved rate of penetration and reduction of non-productive time (Zhang et al., apud Santos et al., 2000; Staysko et al., 2011). Although the benefits of using rotary percussive drilling in mining are remarkable, when the objective is to carry out geotechnical analyses, this type of subsurface investigation still does not meet the needs due to the non-recovery of undisturbed cores.

Despite the current dependence on the activity of describing drill cores to obtain geotechnical parameters, this is characterized as the simplest and least technological way to obtain reliable information on the subsurface, due to the subjectivity of the process.

The application of geophysical methods, in general, is a more technological and interpretive way of obtaining subsurface information, and this becomes extremely necessary when a direct investigation cannot be carried out. With the advent of rotary percussive drilling, a method that does not recover cores, the best way to obtain geotechnical information will be from geophysical profiles.

Due to the high number of geotechnical parameters that correlate directly, indirectly, and inversely, as well as the constant routine of acquiring data from drillings and geophysical profiles, an excellent opportunity arose to use machine learning to compose a tool for predictability and identification of geotechnical parameters.

2 EXPLORATORY DATA ANALYSIS

With the aim to identify behavior trends and evaluate possible correlations between the geophysical profiling data and the information obtained from the geotechnical description in cores, 6.200m of drilling (75 holes) were used distributed in 10 mines located in different regional geological contexts. About 60% of this total drilled is hard rocks (UCS > 10MPa).

In the same way, comparisons were also made for a specific mine (3.500 m in 18 holes) to minimize differences related to the geological context.

When comparing the velocity of the primary (Vp) and secondary (Vs) seismic waves (as shown in Figure 1) with the uniaxial compressive strength obtained in the description of cores (Read & Stacey, 2009) it is possible to notice an overlap of velocity ranges as a function of resistance, that is, in general, a progression of Vp or Vs is not observed with an increase in UCS. This expected increase occurs subtly when compared to a specific mine and in hard rocks. High variation is also observed in weak rocks and even the inverse phenomenon (increase in mean Vp/Vs with decrease in UCS).

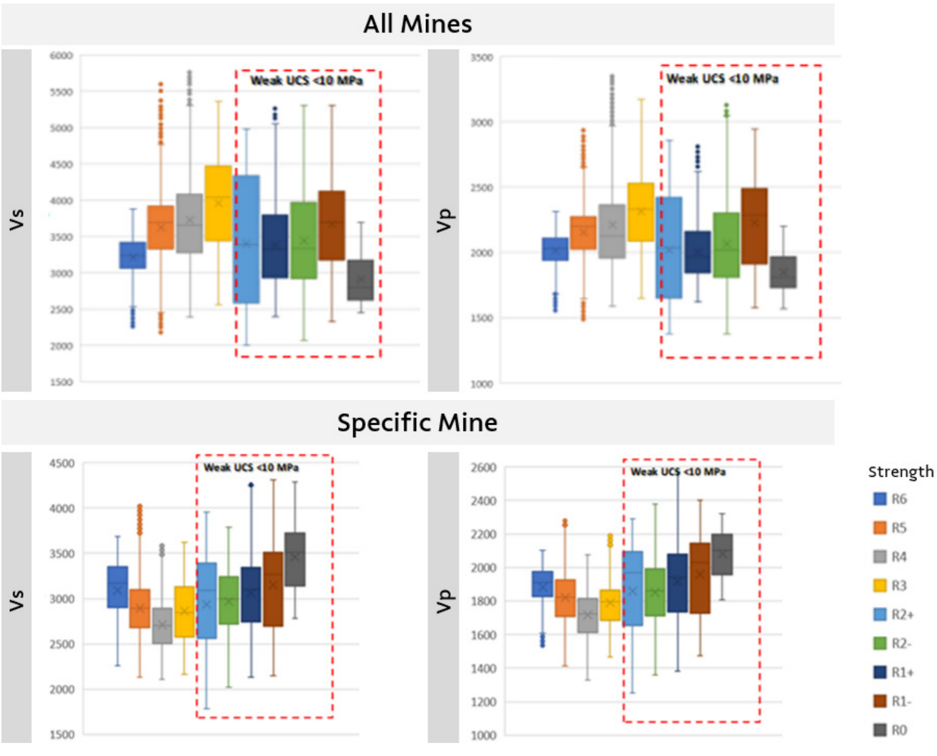

Figure 1. Vp and Vs waves and strength classification.

Comparing Vp with fracturing grade (Figure 2), a high overlap of velocity ranges is observed, but a subtle progression of the mean Vp as a function of the decrease in fracturing grade (ISRM, 1981). This is more evident when compared to a specific mine instead of the unified data of all mines.

Even though there is a high overlapping of Vp variation, when compared with the density of the materials, there is a more evident trend of progression.

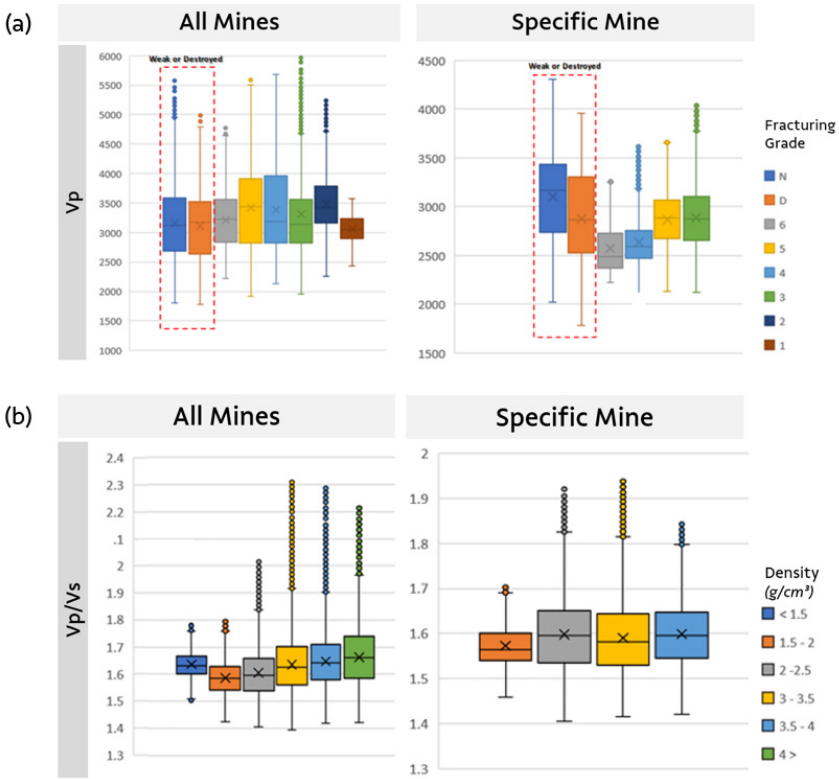

Figure 2. (a) Vp wave and fracturing grade classification (b) Vp/Vs waves and density ranges.

Thus, it was understood that a prediction model between the geophysical logging and the cores description must be built from the comparison of many input data, since the variation of the Vp ranges is superimposed on the variation of the physical parameters of the materials. In addition, the great discrepancy, and inequalities of common sense in the phenomena in which Vp decreases with an increase in geotechnical quality, has a great contribution in the size range (scale) for obtaining each of the methods, while geophysics has a centimeter survey inside the borehole and the cores description has metric scale. It is also noted that the regional geological context greatly affects the analysis, as the most coherent patterns were identified in the comparisons in a specific mine.

2.1 Correlation Matrix

Correlation matrices, according to Friendly (2002) provide the basis for all classical multivariate techniques and can be constructed with the aim of understanding the relationship between database features. In this paper, a correlation matrix was constructed to complement the exploratory data analysis.

The input data for machine learning were Density, Gamma, Depth, Caliper, Vp, Vs and R (strength), among this dataset the methodology must identify which parameters most influence the predictability of another. For this, correlation matrix is built (Figure 4) in which it is possible to identify how each one of the parameters is correlated with the others. The result of this study is presented in the figure bellow.

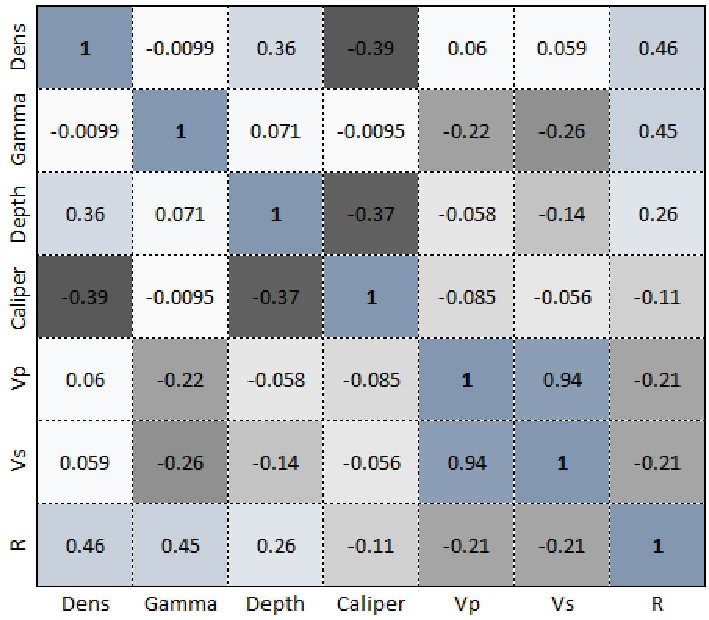

Figure 3. Correlation matrix for strength and geophysical logging parameters.

The correlation matrix indicated that the features that best correlate the predictability of the strength are density (including the Gamma, which is also a parameter for measuring density), followed by depth and velocities of the seismic waves Vp and Vs and Caliper, a tool that measures the preservation/destruction of the walls of a hole.

3 MACHINE LEARNING

Machine learning is used to perform tasks using different algorithms without being explicitly programmed (Mahesh, 2020). It is a concept associated with artificial intelligence, which consists of using mechanisms based on human behavior to solve problems.

Supervised learning is one of several methods of applying machine learning, which consists of inferring a function based on example input-output pairs. Bzdok et al. (2018) define supervised learning algorithms as methods that extract general principles from the observation of examples, guided by a specific prediction goal.

3.1 Geotechnical Parameters Prediction

Applying the data set obtained by geophysical logging to predict the geotechnical core description information, three machine learning techniques were used: Decision Tree, KNN and Neural Network. The results are presented in Table 1.

A decision tree is a classifier consisting of nodes that form a rooted tree. Each tree node divides the sample space into two or more subspaces according to a given function, and the subdivision occurs until the last node, called terminals or decision nodes (Rokach and Maimon, 2005).

The K-Nearest Neighbors (KNN) classification method, in turn, classifies a target from the K closest samples (Kramer, 2013). Based on the idea that the closest patterns provide useful information about a pattern whose class is being searched, the KNN algorithm is, as reported by Guo et al. (2003), a simple but effective classification method.

Artificial neural network algorithms are methods inspired by biological neural networks, which contain nodes that communicate with other nodes through connections (Choi et al., 2020). Among the types of neural networks, the Multilayer Perceptron stands out, an algorithm that simulates signals in one direction and in several layers, from input to output (Popescu et al., 2009).

Table 1. Algorithms scores for datasets.

	All Mines		Specific Mine		Hard Rock	
	Strength Class.	Grade Fracturing	Strength Class.	Grade Fracturing	Strength Class.	Grade Fracturing
Decision Tree	16%	12%	45%	48%	29%	46%
KNN	18%	17%	33%	37%	32%	54%
Neural Network	29%	22%	75%	71%	47%	58%

As for the evaluation, the prediction of the methods in all cases, the Neural Network was the most adequate, reaching a percentage of success between 70 and 75% when applied within the dataset of a specific mine, and up to 58% for hard rocks. The score for the three algorithms when predicting values for all mines together were low (less than 30%).

Figure 4. Confusion matrices for strength prediction.

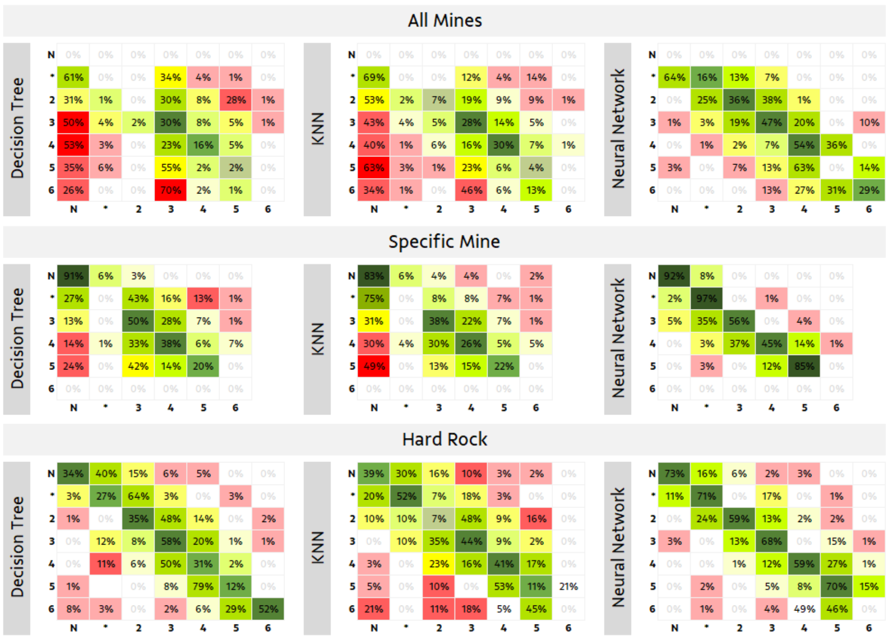

Figure 5. Confusion matrices for fracturing grade prediction.

The difference of scores between datasets can be explained by the lack of physical representation of material parameters that vary in each geological context. Therefore, within the same geological context, the prediction worked reasonably well, but still subject to adjustments and the inclusion of more information obtained from other geophysical logging methods.

For the general set (all mines together) when comparing only the hard materials, there is a 20-30% increase in the accuracy level, that is, twice the accuracy of the compact materials in relation to the complete set. This prediction problem in weak materials was also evident in the construction of the exploratory analysis of the data (Figure 1 and 2a), which is aggravated by the difference in scale of the profiling methods and the geotechnical description.

Another point to be analyzed is the dispersion of errors, as large deviations are extremely harmful. However, small errors can sometimes be tolerated. Furthermore, when increasing the level of acceptability within a new and flexible acceptance criterion (one class below and one above) the level of accuracy rises considerably, reaching levels of predictability that are acceptable to apply in geotechnical project.

4 CONCLUSION

Searching for a more efficient drilling, sometimes methods are chosen that do not recover drill cores, which makes it impossible to survey geotechnical parameters. With the use of geophysical logging, it is possible to recover part of the geotechnical information through the correlation between the physical and geotechnical properties.

This work aimed the utilization of geophysical logging data in machine learning methodologies to predict geotechnical information in drill holes. An exploratory analysis of the data was carried out for a broad understanding of how these data are correlated and then a correlation matrix was set up and the most important parameters for prediction and accuracy were identified.

In general, the prediction shows medium to low accuracy for all algorithms, specifically Decision Tree and KNN. When it is assumed that small sweeps are acceptable the accuracy level goes up considerably for Neural Network, for hard rock and specific mine datasets.

REFERENCES

Bzdok, D., Krzywinski, M., & Altman, N. 2018. *Machine learning: supervised methods*. Nature methods, 15(1), 5. DOI: https://doi.org/10.1038/nmeth.4551

Choi, R., Coyner, A., Kalpathy-Cramer, J., Chiang, M., & Campbell, J. 2020. *Introduction to machine learning, neural networks, and deep learning.* Translational Vision Science & Technology, 9(2), 14-14.

Friendly, M. 2002. *Corrgrams*: Exploratory displays for correlation matrices. The American Statistician, 56(4), 316-324. DOI: https://doi.org/10.1198/000313002533

ISRM. 1981. *Rock characterization. Testing and monitoring.* ISRM Suggested Methods. Ed. Brown, E. T. Commission on testing and monitoring. International Society of Rock Mechanics. Pergamon Press.

Kramer, O. 2013. *K-Nearest Neighbors.* In: Dimensionality Reduction with Unsupervised Nearest Neighbors. Intelligent Systems Reference Library, vol 51. Springer, Berlin, Heidelberg.

Mahesh, B. 2020. Machine learning algorithms-a review. International Journal of Science and Research (IJSR) 9, 381-386.

Popescu, M. C., Balas, V. E., Perescu-Popescu, L., & Mastorakis, N. 2009. *Multilayer perceptron and neural networks*. WSEAS Transactions on Circuits and Systems, 8(7), 579-588.

Read, J. & Stacey, P. 2009. *Guidelines for Open Pit Slope Design.* SCIRO Publishing, Clayton.

Rokach, L., Maimon, O. 2005. *Decision Trees*. In: Maimon, O., Rokach, L. (eds) Data Mining and Knowledge Discovery Handbook. Springer, Boston, MA. https://doi.org/10.1007/0-387-25465-X_9

Xi, Y., Wang, W., Fan, L., Zha, C., Li, J., & Liu, G. 2022. *Experimental and numerical investigations on rock-breaking mechanism of rotary percussion drilling with a single PDC cutter*. Journal of Petroleum Science and Engineering, 208, 109227. DOI: https://doi.org/10.1016/j.petrol.2021.109227

Zhang, X., Zhang, S., Luo, Y., & Wu, D. 2019. *Experimental study and analysis on a fluidic hammer—an innovative rotary-percussion drilling tool*. Journal of Petroleum Science and Engineering, 173, 362-370.