Assessment of mine slopes stability conditions using a decision tree approach

http://dx.doi.org/10.1590/0370-44672021760098

Naim Khalil Ayache[1,3]
https://orcid.org/0000-0003-3834-6341
Allan Erlikhman Medeiros Santos[1,4]
https://orcid.org/0000-0003-4302-3897
Denise de Fátima Santos da Silva[2,5]
https://orcid.org/0000-0002-9695-2449

[1]Centro Federal de Educação Tecnológica - CEFET-MG, Departamento de Minas e Construção Civil, Araxá - Minas Gerais - Brasil.

[2]Universidade Federal de Minas Gerais - UFMG, IGC - Programa de Pós-graduação em Geologia, Belo Horizonte - Minas Gerais - Brasil.

E-mails: [3]naimayache98@gmail.com, [4]allanerlikhman@cefetmg.br, [5]denisefss@ufmg.br

Abstract

Continuous assessment of slope stability is important to the open pit design and operation. This article aims to present a tool for evaluating the stability conditions of rock slopes in mining, based on a global geotechnical database, using machine learning techniques. Different models are evaluated in this research: the general model, which uses all variables; the mathematical model, which uses only variables selected by the random forest (out-of-bag); and two expert-based models: the Q-Slope model and the Santos model. The validation of the model was done through the test sample, using partition confusion matrices aiming at reproducibility of the results. A study of the types of errors was carried out using Principal Component Analysis (PCA). The study of errors allowed the identification of samples that were inconsistent with the others. Afterwards, the models were redone and compared with the previous ones. The best performers are presented and discussed. The proposed methodology does not replace the classic analysis of slope stability. On the contrary, it contributes to engineers and geologists with a tool for monitoring the stability conditions of slopes in a mining operation. Slope stability analysis must be carried out throughout the mine's lifetime and, therefore, it is believed that the tool proposed here can optimize the selection of slopes most susceptible to instability.

keywords: stability conditions of mine slopes, decision trees, random forest, geomechanical parameters.

1. Introduction

Open-pit mining is the most important technology for extracting mineral resources from the earth's crust (Zare Naghadehi *et al*., 2013). Mining slopes are applied in all phases of the open-pit mining project, mine development, mining exploitation and mine closure. Slopes allow access conditions to the geological body at different levels, with operational safety, e.g., transport and excavation.

The instability on slopes is conditioned by the geological-geotechnical characteristics of the rock mass, the presence of discontinuities, applied geometry and external factors. The occurrence of rupture impacts the mining activities, especially the operational, economic, and environmental sectors.

The aim of the research is the study of empirical models to assess the stability of mine slopes using machine learning techniques. Models capable of interpreting information taken from slope databases and generating a reliable estimate of the rock mass stability conditions are presented. The database used was proposed by Zare Naghadehi *et al.* (2013).

The authors present the development of a new Mine Slope Instability Index (MSII) which aims to determine the stability conditions of mining slopes in open pit operations, using artificial neural networks and the RES system proposed by Hudson (1992). The database has been used in recent research, with promising results and accepted by the technical community, e.g., Santos *et al.* (2018), Santos *et al.* (2019).

The general methodology was based on the construction of different models: a general model with all the variables of the database; a mathematical model with variables selected from their importance, using Random Forest, to choose the variables in order to target slope stability; two expert models using variables applied in classification systems, based on Q-Slope (Bar & Barton, 2017) and Santos *et al.* (2021). The model validation was done through the test sample, using bootstrap and partition confusion matrices aiming at the reproducibility of the results. A study of the errors using Principal Component Analysis (PCA) allowed the identification of samples inconsistent with the others, so the models were remade and compared with the previous ones. This way the best modeling was found based on the variables selected by Random Forest with the database without the problematic samples.

The technique applied herein is similar to the study presented by Santos *et al.* (2018), that is, decision trees applied to a similar dataset, but being different in the proposed models, specifically the set of variables that make up each model presented. In this study, four models are proposed: the general model with all the variables in the database, the mathematical model in which the variables were selected by the Gini index of Random Forest, measuring the importance of each variable for the classification problem. The model with variables based on the Q-Slope (Bar and Barton, 2017). And finally, a model based on the variables proposed by Santos *et al*. (2021). Among the models presented, the General Model is the closest to Santos *et al.* (2018), since it uses all variables from the database proposed by Zare Naghadehi, *et al.* (2013).

The application of machine learning techniques has been increasing in mining, with positive impacts, in recent years. In addition to the studies already mentioned, some studies with machine learning applications can be highlighted are, e.g., Klen & Lana (2014), Silva *et al*. (2018), Baretta *et al*. (2019), Okada *et al*. (2019), Santos *et al*. (2020). The methodology presented allows different users, target audiences in general, to apply the model quickly and accurately, optimizing decisions in mining operations.

2. Material and methods

2.1 Initial considerations

The methodology was developed in freeware R (R Core Team, 2016). The methodology was applied to the database compiled and organized by Zare Naghadehi *et al*. (2013). The database organized by Zare Naghadehi *et al*. (2013) presents 84 samples with 18 predictive variables, from mines around the world.

Zare Naghadhi *et al*. (2013) in their article classified the slopes studied into three distinct categories: the first are the stable slopes (ST), the second are the unstable slopes (OF) and finally the slopes with punctual bench failures (FSB).

In this study, models of classification decision trees were developed, which were named General Model (MG), Mathematical Model (MM), Q-slope Model (MQ), Santos Model (MS), Mathematical Models without errors (MMS) and finally Model Santos without errors (MSS). Each model has its characteristics that will be described in the text below.

2.2 Pre-processing

The main pre-processing issue for this database was the imbalance classes. According to Vladislavleva *et al.* (2010) in predictive models, class balancing is essential to avoid classifier bias. Every classifier has a weak a priori classifier and this is influenced by the distribution of classes. Therefore, an imbalance can negatively influence the training of the model.

The class balancing process was developed with the ROSE package (Lunardon *et al*.2014). The Random Over-Sampling Examples (ROSE) provide functions for dealing with the binary classification of problems in the presence of unbalanced classes. Synthetic balanced samples are generated according to ROSE.

2.3 General methodology

After pre-processing, the methodology was applied. The models MG, MM, MQ and MS were trained. For these models the packages were used, and the first was the rpart (Therneau & Atkinson, 2019) and then, the partykit (Hothorn & Zeileis 2015). Figure 1 presents the general research flowchart.

Recursive Partitioning and Regression Trees (rpart) is a package aimed at recursive partitioning for the classification, regression trees and classification, using the concepts implemented in the article of Breiman *et al.* (1984). *The Toolkit for Recursive Partytioning* (partykit) is a set of tools with functions to represent, summarize, and visualize tree-structured regression and classification models.

Incorrectly classified samples were studied, and consequently, two new models were proposed, the MMS and MSS. The errors study was implemented by Principal Component Analysis (PCA). Figure 2 presents the study's error flowchart.

For the creation of the models, the balanced data were divided into two distinct subsets, the Training and Test Sets. The Training Set had 80% of balanced data and was used to create the models, serving as a basis for learning the algorithms. The Test Set had the remaining 20% of the data balanced and was used to validate the models.

2.4 Description of models

In the General Model (MG), all 18 predictive variables from the database developed by Zare Naghadehi *et al*. (2013) were used to estimate the slope stability variable. In the Mathematical Model (MM), a selection of variables was previously performed using the Out-of-bag (OOB) of Random Forest (RF), in the same database, to assign weights to the variables with greater importance for determining the objective variable of stability. For variable selection in MM, from OOB-RF, the varSelRF package, proposed by Diaz-Uriarte and Alvarez de Andres (2005), was applied.

The Q-Slope Model (MQ) was based on the studies by Bar and Barton (2017) to create a relationship between the variables from the authors' empirical model (Q-Slope) with the variables from the database. The use of this model had as main objective, to compare the effect of the variables of the empirical model in the methodology of the decision trees and to cross these data with the other modeling.

The Santos Model (MS) was developed based on the studies by Santos *et al.* (2021), a proposal to improve the RMR classification system, proposed by Bieniawski (1989). The techniques applied in the research by Santos *et al.* (2021) come from the areas of multivariate statistics and artificial intelligence. According to the identification of geomechanical variables common to the RMR, Santos *et al.* (2021) determined three different factors to determine the quality of the rock mass. These same variables were selected for the development of MS.

2.5 Study of errors, MMS and MSS models

The MG, MM, MQ and MS models were validated using the training/test sets. The database visualization in 2-dimensions (applied PCA) allowed to visualize the error frequency in specific samples.

Samples that were misclassified were often removed from the original database and the entire sample preparation procedure was redone in a new database (without the samples that were always misclassified). The two models with the best results, among the MG, MM, MQ and MS, were selected to be developed again with the new database. In this case, "the best result" is the one considered "dangerous error", that is, classification errors in which the estimate presented a stability superior to the real one of the dataset.

The discussion about the removal of these samples is aimed at mainly studying the effect they had on the final accuracy of the models. As each model had different variables in its composition, it was expected that there would also be wrong estimates in different samples. Therefore, the recurrence of wrongly classified samples motivated this approach, as can be seen later.

Using the same procedures as the MM and MS, the Error-free Mathematical (MMS) and Error-free Santos (MSS) Models were developed with the new balanced Test Set and validated with the new balanced Training Set. The obtained accuracy results for all models were compared to determine the best final decision tree models obtained.

2.6 Repository codes for reproducing the applied methodology

The repository codes can be found in GitHub, see the link:
https://github.com/MrColugo/Slope-stability-study-with-Decision-Tree-and-Random-Forests-.git

3. Results and discussions

The balancing process was applied to the database by Zare Naghedehi *et al.* (2013). In Figure 1, it is possible to observe the distribution of unbalanced classes (a) and after the balancing procedure (b).

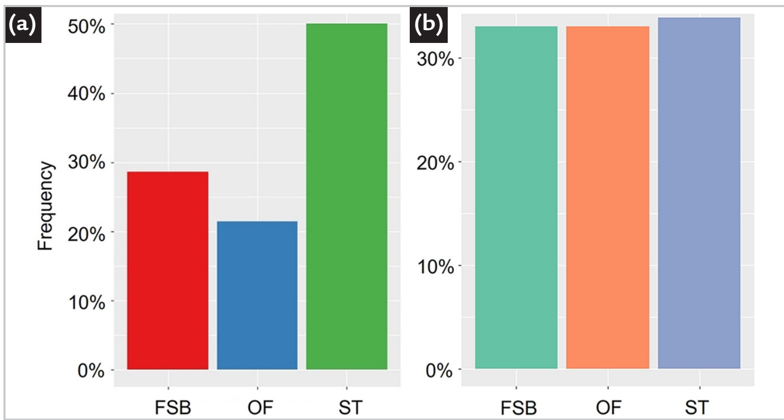

Figure 1 - Distribution of unbalanced classes (a) and after the balancing procedure (b).

In the MG model, all variables from the database by Zare Naghadehi *et al.* (2013), were used. The MM model consists of the variables selected by the OOB of Random Forest. MQ was based on the components of the Q-slope equation. The MS model was based on the factors of Santos *et al.* (2021). Table 1 shows the variables used in the MM, MQ and MS models.

Table 1 – Variables used for models MM, MQ and MS.

Model	Variables used of Zare Naghadehi *et al.* (2013)
General Model (MG)	All variables are used.
Mathematical Model (MM)	Intact rock strength; RQD; Weathering; Discontinuity families; Discontinuity Orientation; Slope angle; Previous instability.
Q-slope model (MQ)	Intact rock strength; RQD; Weathering; Discontinuity families; Discontinuity spacing; Discontinuity Orientation; Discontinuity aperture; JRC; Discontinuity filling; Blasting method; Precipitation.
Santos Model (MS)	Intact rock strength; Weathering; Groundwater water; Discontinuity persistence; Discontinuity spacing; Discontinuity aperture.

To test the models, the test set consisted of 20% of the database, which was submitted to the interpretation of each decision tree. Table 2 presents the general accuracy of each model with their respective Kappa indices, which means the reproducibility of the results. The accuracy values are within ranges that validate the application of decision trees and, consequently, the selection of variables.

Table 2 – Accuracy and Kappa for models MM, MQ and MS.

Metric	MG	MM	MQ	MS
Accuracy	0.8333	0.8333	0.7917	0.8333
Kappa index	0.7500	0.7500	0.6875	0.7500

Tables 3 to 6 present the by class statistics for the models. Sensitivity, efficiency and balanced accuracy metrics showed a good fit for a multiclass problem; no bias is verified in the predictive models.

Table 3 – Assessment Metrics for MG model.

MG model	FSB	OF	ST
Sensitivity	0.8571	1.0000	0.7273
Specificity	0.8824	0.8889	1.0000
Balanced accuracy	0.8697	0.9444	0.8636

Table 4 – Assessment Metrics for MM model.

MG model	FSB	OF	ST
Sensitivity	0.8333	1.0000	0.7273
Specificity	0.8333	0.9412	1.0000
Balanced accuracy	0.8333	0.9706	0.8636

Table 5 – Assessment Metrics for MQ model.

MG model	FSB	OF	ST
Sensitivity	0.6667	0.875	0.8571
Specificity	0.8667	0.9375	0.8824
Balanced accuracy	0.7667	0.9062	0.8697

Table 6 – Assessment Metrics for MS model.

MG model	FSB	OF	ST
Sensitivity	0.6667	1.0000	1.0000
Specificity	1.0000	0.9412	0.8421
Balanced accuracy	0.8333	0.9706	0.9211

After training the models, all database (training and test samples) were applied to the trees and plotted in a reduced space (2-dimension, by PCA). The objective of this was to verify the confusion zones and verify the error frequency for each sample. Figures 2 to 5 present the result for each model.

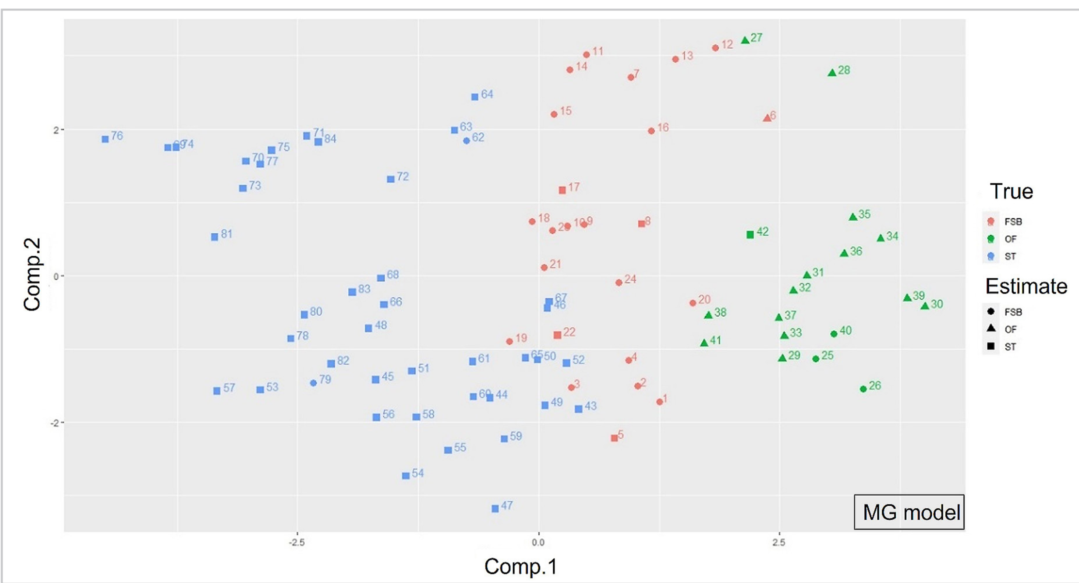

Figure 2 – Study of the error, in 2-dimension, for MG model.

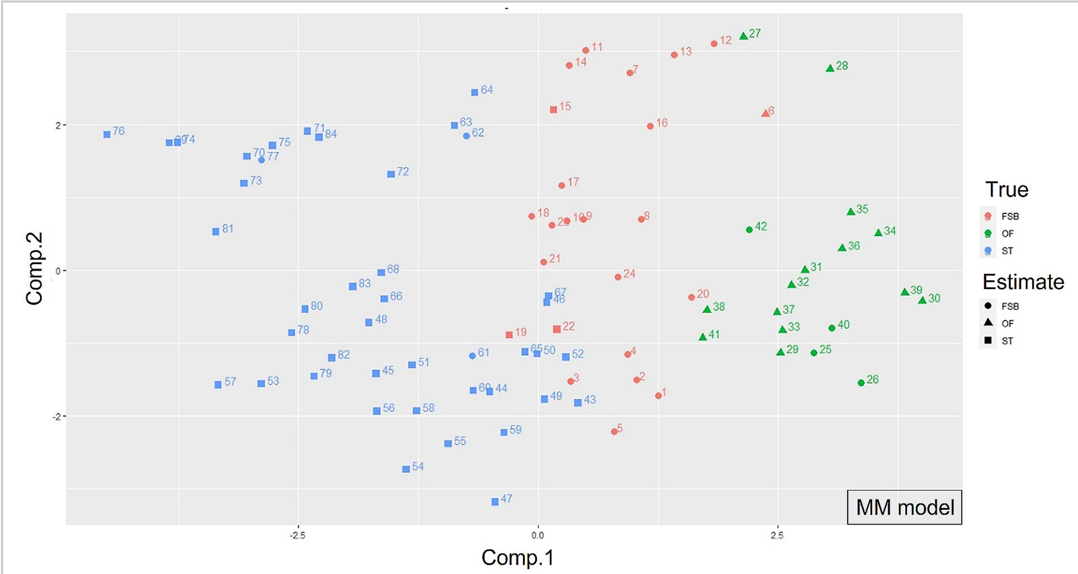

Figure 3 – Study of the error, in 2-dimension, for MM model.

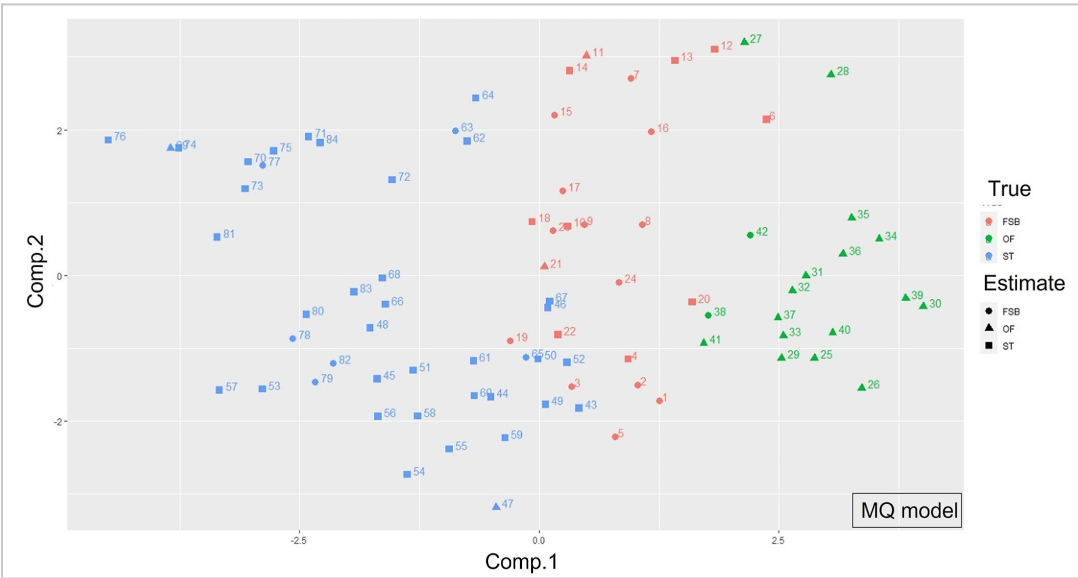

Figure 4 – Study of the error, in 2-dimension, for MQ model.

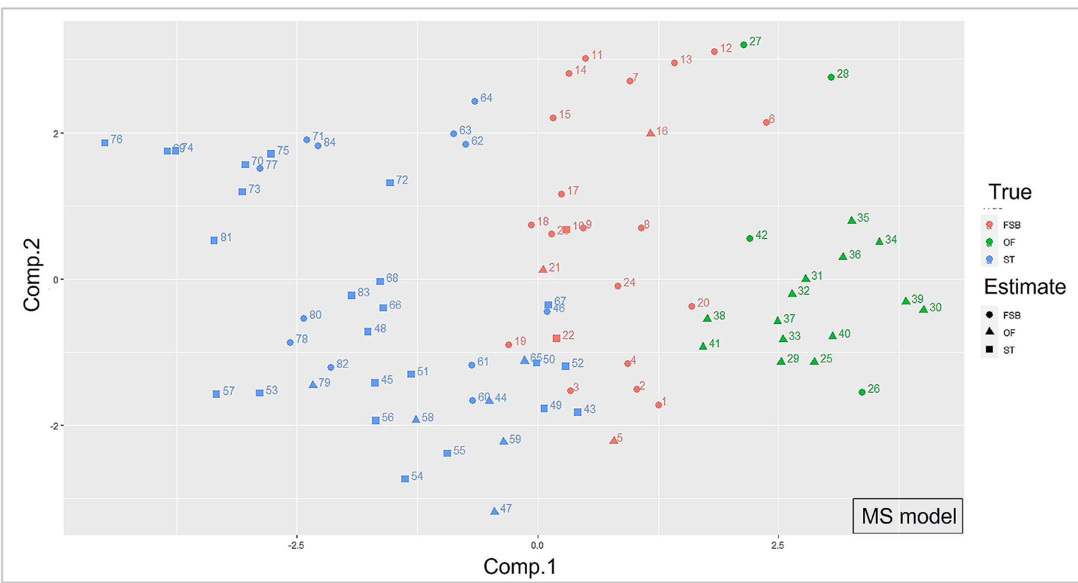

Figure 5 – Study of the error, in 2-dimension, for MS model.

From Figures 2 to 5, it was possible to assemble Table 7 that presents the samples by id that were incorrectly classified for each model. The red highlighted id's in Table 7 are the "dangerous errors". Therefore, the frequency of errors was obtained.

Table 7 - List of incorrectly estimated samples for each model.

Model	Id Sample	Total errors	"Dangerous Error" proportion
MG	5, 6, 8, 17, 21, 22, 25, 26, 42, 62, 79	11	0.63
MM	6, 21, 22, 25, 26, 40, 42, 71, 77, 82, 84	11	0.54
MQ	4, 6, 10, 11, 12, 13, 14, 18, 20, 21, 22, 38, 42, 47, 63, 65, 69, 77, 78, 79, 82	21	0.52
MS	5, 10, 16, 21, 22, 26, 27, 28, 42, 44, 46, 47, 58, 59, 60, 61, 62, 63, 64, 65, 71, 77, 78, 79, 80, 82, 84	27	0.22
Repeated	5, 6, 21, 22, 25, 26, 42, 10, 16, 47, 65, 71, 77, 79, 82, 84	16	

The error study shows that, although the MG and MM models have a lower error rate, the concentration of dangerous errors is high. The MG models use all the variables, and the MM is optimized with a focus on math metrics only. Although MS has a higher error rate, the concentration of dangerous errors is low, 0.22, which can be interpreted as a good selection of variables performed by Santos *et al.* (2020). This result reinforces the importance and impact of variable selection for predictive models.

New MMS and MSS models were trained using the result of the error study. The chosen models were those that presented the lowest number of "dangerous errors", that is, the MM and the MS. Samples that were misclassified in all models (repeatedly), presented in row 6 of Table 7, were removed from the database and a new database was used to build the MMS and MSS models. The same procedures as the MM and MS were applied to modeling the MMS e MSS. The Figures 6 and 7 presents de decision tree for each model.

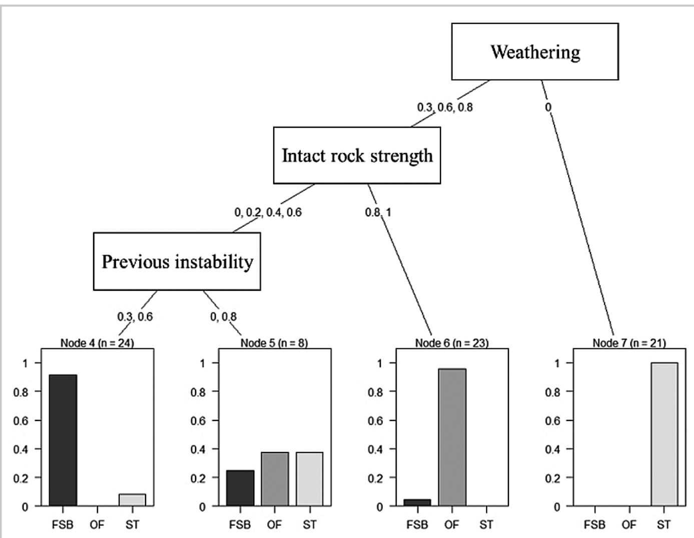

Figure 6 – Decision tree for MMS model.

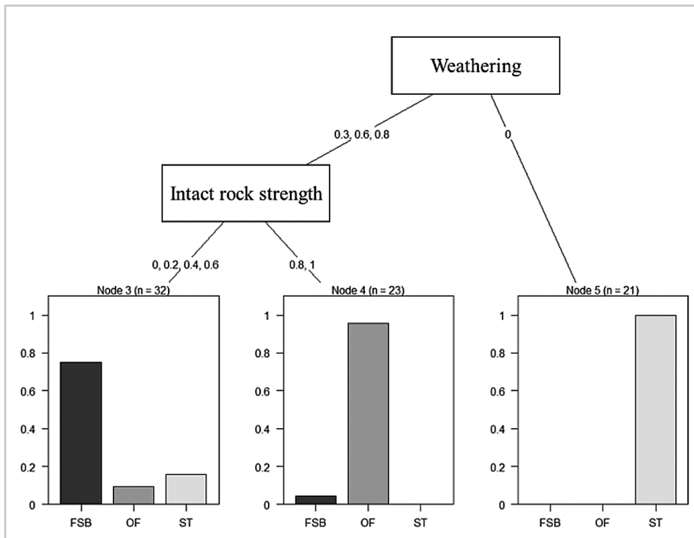

Figure 7 – Decision tree for MSS model.

Table 8 presents the confusion matrix for the test sample applied to the MMS and MSS models. Note that the models presented only 1 error, the MSS error is a dangerous error, an OF sample is classified as FSB.

Table 8 - Confusion Matrix to MMS and MSS.

	Predicted					
	MMS			MSS		
True	FSB	OF	ST	FSB	OF	ST
FSB	5	1	0	6	0	0
OF	0	6	0	1	5	0
ST	0	0	6	0	0	6

The accuracy and Kappa index of the MMS and MSS models were equal, with value 0.94 to accuracy and 0.91 to Kappa index. The evaluation metrics by class are shown in Tables 9 and 10. The results show the improvement of the results presented by the MS and MM models, previously shown in Tables 2, 4 and 6. This result is expected as inconsistent samples were removed, causing a database cleanup.

Table 9 – Assessment Metrics for MMS model.

MMS model	FSB	OF	ST
Sensitivity	1.0000	0.8571	1.0000
Specificity	0.9231	1.0000	1.0000
Balanced accuracy	0.9615	0.9286	1.0000

Table 10 – Assessment Metrics for MSS model.

MSS model	FSB	OF	ST
Sensitivity	0.8571	1.0000	1.0000
Specificity	1.0000	0.9231	1.0000
Balanced accuracy	0.9286	0.9615	1.0000

4. Conclusions

By using models such as decision trees, it is possible to create a direct and simple way to interpret the stability conditions of a slope. This research has as its main objective the proposal of creating models that are easy to implement and with a satisfactory efficiency in the field to determine the slope stability. From a database with 84 samples collected from slopes around the world, 6 decision tree models were created, using different variables from mathematical and literature interpretations of different geotechnical and spatial parameters of slopes.

After performing all stages of development, it was possible to determine that the best models were the Mathematical Model and the Santos Model. In addition to having a high accuracy mainly for a simple model, such as decision trees. Aa low incidence of dangerous errors was also obtained, which further increases their potential use for slope stability estimates.

With the use of these developed models, it is possible to determine the stability conditions of slopes of open-pit mines on an industrial scale, being able to vary the different geotechnical parameters to evaluate the result of the interaction of the variables. Mainly in the control of the height and general angle of the slopes, which are essential to determine the progress of mineral activities in any open-pit mining project. In this way, it is possible to optimize the exploitation and use of the reserve, keeping the pit operational and maximizing the safety of operations.

Furthermore, as they use variables easily obtained in the field, these models can be used by users in general. The models presented here deconstruct the "Black-boxes" present in artificial intelligence models that limit the use of a general public. This facilitates decision-making in projects involving these types of problems. Finally, as a proposal for future work, there is the possibility of adding new data sets in order to refine the prediction of the models.

Acknowledgements

The authors thank CEFET-MG, UFMG and CIDENG CNPq for their support during the research.

References

BAR, N.; BARTON, N. The Q-Slope method for rock slope engineering. ***Rock Mechanics and Rock Engineering***, v. 50, n. 12, p. 3307-3322, 2017. Disponível em: http://dx.doi.org/10.1007/s00603-017-1305-0.

BERETTA, F.; RODRIGUES, Á. L.; PERONI, R. de L.; COSTA, J. F. C. L. Using UAV for automatic lithological classification of open pit mining front. ***REM - International Engineering Journal***, v. 72, n. 1 Supl. 1, p. 17-23, 2019. Disponível em: https://doi.org/10.1590/0370-44672018720122.

BIENIAWSKI, Z. T. ***Engineering rock mass classifications***: a complete manual for engineers and geologists in mining, civil, and petroleum engineering. Wiley-Interscience, 1989. 272p.

BREIMAN, L.; FRIEDMAN, J. H.; OLSHEN, R. A.; STONE, C. J. Classification and regression trees. Routledge, 1984. 368p. Disponível em: https://doi.org/10.1201/9781315139470.

DIAZ-URIARTE, R.; DE ANDRÉS, S. A. **Variable selection from random forests**: application to gene expression data. arxiv 2005. Disponível em: https://arxiv.org/pdf/q-bio/0503025.pdf. Accesso em: 19 July 2021.

HOTHORN T, ZEILEIS, A. Partykit: a modular toolkit for recursive partytioning in ***R. Journal of Machine Learning Research***, v. 16, p. 3905–3909, 2015. Disponível em: http://jmlr.org/papers/v16/hothorn15a.html.

HUDSON, J. A. ***Rock engineering systems*, theory and practice**. Chichester: Ellis Horwood, 1992. 185p.

KLEN, A. M.; LANA, M. S. Fuzzy algorithm of discontinuity sets. ***REM - International Engineering Journal***, v. 67,

n. 4, p 439-445, 2014.

LUNARDON, N.; MENARDI, G.; TORELLI, N. ***R package ROSE***: random over-sampling examples (version 0.0-3). Università di Trieste and Università di Padova, Italia, 2013. Disponível em: http://cran.r-project.org/web/packages/ROSE/index.html. [p79]

OKADA, R.; COSTA, J. F. C. L.; RODRIGUES, Á. L.; KUCKARTZ, B. T.; MARQUES, D. M. Scenario reduction using machine learning techniques applied to conditional geostatistical simulation. ***REM - International Engineering Journal***, v. 72, n. 1 Suppl 1, p. 63–68, 2019. Disponível em: https://doi.org/10.1590/0370-44672018720135

TEAM. R. C. R: a language and environment for statistical computing. ***R Foundation for Statistical Computing***, Vienna, Austria. 2016. Disponível em: https://www.R-project.org/.

SANTOS, A. E. M.; AMARAL, T. K. M.; MENDONÇA, G. A.; SILVA, D. F. S. Open stope stability assessment through artificial intelligence. ***REM - International Engineering Journal***, v. 73, n. 3, p. 395-401, 2020.

SANTOS, A. E. M.; LANA, M. S.; CABRAL, I. E.; PEREIRA, T. M.; ZARE NAGHADEHI, M.; SILVA, D. F. S.; SANTOS, T. B. Evaluation of rock slope stability conditions through discriminant analysis. ***Geotechnical and Geological Engineering***, v. 37, p. 775–802, 2019. Disponível em: https://doi.org/10.1007/s10706-018-0649-x

SANTOS, A. E. M.; LANA, M. S.; PEREIRA, T. M. Evaluation of machine learning methods for rock mass classification. ***Neural Computing and Applications,*** 2021. Disponível em: https://doi.org/10.1007/s00521-021-06618-y

SANTOS, A. E. M.; LANA, M. S.; PEREIRA, T. M. Rock mass classification by multivariate statistical techniques and artificial intelligence. ***Geotechnical and Geological Engineering,*** v. 39, p. 2409–2430, 2021. Disponível em: https://doi.org/10.1007/s10706-020-01635-5

SANTOS, T. B. dos; LANA, M. S.; PEREIRA, T. M.; CANBULAT, I. Quantitative hazard assessment system (Has-Q) for open pit mine slopes. ***International Journal of Mining Science and Technology***, v. 29, n. 3, p. 419-427, 2018. Disponível em: https://doi.org/10.1016/j.ijmst.2018.11.005

SILVA, D. F. S.; SANTOS, A. E. M.; FERREIRA, B. T.; PEREIRA, T. M.; CORTELETTI, R. C. Cluster analysis for slope geotechnical prioritization of intervention for the Estrada de Ferro Vitória - Minas. ***REM - International Engineering Journal***, v. 71, n. 2, p. 167-173, 2018. Disponível em: https://doi.org/10.1590/0370-44672017710173.

THERNEAU, T.; ATKINSON, B. **Rpart: recursive partitioning and regression trees**. 2019. Disponível em: https://cran.r-project.org/package=rpart

VLADISLAVLEVA, E.; SMITS, G.; DEN HERTOG, D. On the importance of data balancing for symbolic regression, ***IEEE Transactions on Evolutionary Computation***, v. 14, n. 2, p. 252-277, 2010. DOI: 10.1109/TEVC.2009.2029697.

ZARE NAGHADEHI, M.; JIMENEZ, R.; KHALOKAKAIE, R.; ESMAEIL JALALI, S. A new open pit mine slope instability index defined using the improved rock engineering systems approach. ***International Journal of Rock Mechanics and Mining Sciences***, v. 61, p. 1-14, 2013. Disponível em: https://doi.org/10.1016/j.ijrmms.2013.01.012

Received: 6 December 2021 - Accepted: 5 July 2022.