

Análise de Importância de Variáveis na Diluição de Realces Subterrâneos por Meio de Técnicas de Aprendizado de Máquina

Caio Oliveira Rodrigues

Engenheiro de minas e geotécnico, Universidade Federal de Ouro Preto (UFOP), Ouro Preto, Brasil, caio.or@aluno.ufop.edu.br

Tatiana Barreto dos Santos

Professora Adjunta do Departamento de Engenharia de Minas (DEMIN/UFOP), Universidade Federal de Ouro Preto, Ouro Preto, Brasil, tatiana.santos@ufop.edu.br

RESUMO: A partir de um banco de dados com 70 realces de três diferentes mineralizações (Carvoaria, Cachorro Bravo e Laranjeiras) e 25 variáveis locais, geométricas, geológico-geotécnicas e de perfuração e desmonte, foram conduzidas análises por métodos de aprendizado de máquina, a fim de determinar as variáveis mais relevantes na ocorrência de diluição operacional da mina em estudo. Os métodos escolhidos foram Árvores de Decisão (DT) e Florestas Aleatórias (RF), sendo este capaz de determinar a importância de variáveis por meio do estimador de erros Out-of-Bag (OOB). O estimador OOB foi mais eficiente na definição de significância de variáveis, com valores de R^2 superiores aos obtidos por meio das DT. Os resultados do corpo Carvoaria indicaram que adequações nas geometrias de projeto dos realces são demandadas. Para o corpo Cachorro Bravo há necessidade de atenções e cuidados direcionados às tensões *in-situ* e induzidas, enquanto para o corpo Laranjeiras as análises sugerem que o plano de fogo deve ser redimensionado, para evitar rupturas por quedas de blocos que causam diluição não planejada. Apesar das restrições de amostras disponíveis no banco de dados, o trabalho conseguiu mostrar a relevância de variáveis na diluição de cada mineralização e propor melhorias direcionadas para cada caso.

PALAVRAS-CHAVE: Diluição, Realces por Subníveis, Árvores de Decisão, Florestas Aleatórias, Estimador de Erros *Out-of-Bag*.

ABSTRACT: From a dataset with 70 stopes of three different orebodies (Carvoaria, Cachorro Bravo, and Laranjeiras) and 25 locational, geometrical, geological-geotechnical, and drilling and blasting variables. Machine learning analyses are conducted to determine the most relevant variables that cause stope dilution. The chosen methods were Decision Trees (DT) and Random Forests (RF), the latter capable of determining variable importance through the Out-of-Bag (OOB) error estimator. OOB estimator was more efficient in defining variable significance, with R^2 values higher than those obtained through DT. The results for the Carvoaria orebody indicate that adjustments to the design geometries of stopes are required. For the Cachorro Bravo orebody, attention are needed for *in-situ* and induced stresses, while for the Laranjeiras orebody, the analyses suggest that blasting plan should be resized to avoid block falls that cause unplanned dilution. Despite the sample limitations in the database, the paper manages to show the relevance of variables in each dilution mineralization and propose targeted improvements for each case.

KEYWORDS: Dilution, Sublevel Stopping, Decision Trees, Random Forests, Out-of-Bag Error.

1 INTRODUÇÃO

Um dos principais desafios na mineração subterrânea pelo método de lavra por realces por subníveis é o controle da diluição operacional. Alguns dos seus impactos diretos são, conforme Clark (1998), Charbel (2015), Jang et al. (2015) e Costa (2017), nos custos de produção, na recuperação metalúrgica praticada na planta de beneficiamento e na diminuição dos teores de minério, que exigem mais operações nas frentes de lavra, diminuindo a produtividade da operação, além de afetar a segurança de trabalhadores e máquinas quando ocorrem altas taxas de diluição.

Costa (2017) afirma que uma compreensão adequada dos fatores que levam à diluição é de suma importância para a continuidade de um empreendimento mineiro. Diversos métodos empíricos foram desenvolvidos historicamente para este fim, e mais recentemente, a utilização de ferramentas de *data mining* têm sido propostas para minimizar a diluição e compreender os fatores associados à sua ocorrência, uma vez que são de utilização prática, reduzem subjetividades de modelos empíricos, possuem boa acurácia e são uma excelente alternativa aos geotécnicos no auxílio de tomadas de decisão. Assim, este trabalho busca analisar por meio de duas técnicas diferentes de aprendizado de máquina (Árvores de Decisão e Florestas Aleatórias) a importância das diversas variáveis que governam a estabilidade de escavações subterrâneas e seu impacto na diluição operacional em uma mina que utiliza o método de realces por subníveis.

2 REFERENCIAL TEÓRICO

2.1 Árvores de Decisão

As árvores de decisão (do inglês *Decision Trees*, DT) são modelos de aprendizado de máquina utilizados em problemas de classificação e regressão. Elas funcionam criando uma estrutura hierárquica de nós internos, galhos e nós folha para representar regras de decisão baseadas nos dados de entrada (SANTOS, 2019). Conforme Qi *et al.* (2018) e Santos (2021), os nós internos representam as variáveis do conjunto de dados, enquanto os galhos são as decisões tomadas com base nessas variáveis. Cada galho leva a um próximo nó interno, refletindo as escolhas feitas pelo modelo. Os nós folha representam os resultados finais da árvore, sendo rotulados com classes (em árvores de classificação) ou valores numéricos (em árvores de regressão).

Lemos *et al.* (2005) afirmam que uma das características mais importantes das árvores de decisão é a sua capacidade de criar regras de classificação do tipo "se-então" para prever classes ou valores de saída. Essas regras são formadas ao longo dos galhos que conectam a raiz aos nós folha, representando combinações de decisões. Além disso, os autores atestam que a estrutura hierárquica das árvores de decisão reflete a relação de dependência entre as variáveis de entrada e a variável de saída, permitindo identificar os atributos mais influentes na previsão. Isso proporciona uma interpretação clara do modelo e dos fatores que mais impactam a variável de saída.

2.2 Florestas Aleatórias

Conforme apresentam Breiman (2001), Santos (2019) e Guedes (2020), as florestas aleatórias (do inglês *Random Forests*, RF) são um método de aprendizado de máquina proposto por Breiman (2001) que combina os conceitos de árvores de decisão e métodos *ensemble*. O método funciona construindo um conjunto de árvores de decisão de forma aleatória, utilizando diferentes amostras geradas pela técnica de *bagging*, onde são selecionados aleatoriamente diversos subconjuntos de treinamento do conjunto de dados original, gerando assim amostras aleatórias para treinar cada árvore.

Cada árvore de decisão na floresta aleatória é criada a partir de uma amostra aleatória independente dos dados de treinamento, o que torna as árvores diversificadas e menos propensas a *overfitting*. Durante a construção de cada árvore, em cada nó de decisão é feita uma seleção aleatória de um conjunto de atributos para avaliar qual é o melhor atributo para dividir o nó. Após a construção das árvores, o resultado final da floresta aleatória é determinado pela votação majoritária das árvores em problemas de classificação, onde a classe mais votada é escolhida como resultado final. Em problemas de regressão, o resultado final é obtido pela média dos valores previstos por todas as árvores (BREIMAN, 2001; OSHIRO, 2013; SANTOS, 2021). Uma das principais vantagens das RF é a sua capacidade de lidar com conjuntos de dados complexos de forma não-linear, além de serem menos sensíveis a *outliers* e ruídos nos dados, e geralmente fornecem resultados de alta precisão em comparação com outros métodos de aprendizado de máquina. (QI *et al.*, 2018)

2.2.1 Estimador de Erros Out-of-Bag

Breiman (2001) justifica o uso do método *bagging* nas florestas aleatórias uma vez que melhora o desempenho do algoritmo, especialmente quando se trabalha com atributos aleatórios. Além disso, ele oferece

uma maneira eficaz de estimar o erro de generalização do conjunto combinado de árvores por meio do estimador *out-of-bag* (OOB), o que proporciona uma avaliação contínua da performance do modelo.

São conhecidos como *out-of-bag* (OOB) os dados não utilizados na construção do modelo, e possuem grande relevância no entendimento da importância das variáveis no contexto das florestas aleatórias. A não utilização frequente de certas variáveis OOB em diferentes iterações das árvores indica sua baixa influência na predição desejada. Isso é uma consequência natural do modelo, pois as variáveis mais relevantes tendem a ser escolhidas repetidamente para a construção das árvores. A estimativa de erro OOB oferece uma medida objetiva do desempenho do modelo, pois é obtida a partir da avaliação das amostras OOB presentes no modelo final. Além disso, ao utilizar o estimador de erro OOB, é possível determinar a importância preditiva de cada variável no modelo. Quanto maior a diferença no erro de predição ao substituir uma variável por variáveis OOB, mais relevante é aquela variável para o modelo preditivo, provendo importantes conclusões sobre a contribuição de cada variável para a precisão do modelo (BREIMAN, 2001; OSHIRO, 2013; SANTOS, 2021).

3 MATERIAIS E MÉTODOS

3.1 Banco de dados

O banco de dados utilizado neste trabalho foi obtido em Costa (2017). O banco de dados inicial é composto por 77 realces, contendo 39 variáveis em cada realce.

Numa análise inicial foram removidas amostras onde figuravam valores nulos ou incoerentes, e também descartadas algumas variáveis para análise posterior por não possuírem relação direta com diluição ou apresentarem muitos valores nulos ou pouco variáveis. Assim, o banco de dados utilizado no trabalho contém 70 realces e 26 variáveis, divididas em quatro grupos de parâmetros apresentados a seguir. Os 70 realces pertencem a três diferentes mineralizações, sendo 25 realces pertencentes ao corpo Carvoeira, 22 ao corpo Cachorro Bravo e 23 ao corpo Laranjeiras. Em parênteses é apresentado a sigla da variável que irá constar nos outros tópicos do trabalho:

- Parâmetros locais: profundidade (Prof, em metros);
- Parâmetros geométricos/de projeto: comprimento do realce (Strike, em metros), altura do realce em sua inclinação (AltDip, em metros), ângulo de mergulho do *hanging wall* do realce (HwDip, em graus), raio hidráulico do *hanging wall* (RH, em metros), espessura planejada do realce (Ep, em metros), largura da galeria inferior ao realce (LGi, em metros), largura do *undercutting* na galeria inferior ao realce do lado do *hanging wall* (LUGIhw, em metros) e largura do *undercutting* na galeria inferior ao realce do lado do *footwall*, em metros (LUGIfw, em metros);
- Parâmetros geológico-geotécnicos: classificações geomecânicas Q' , RMR e GSI, fatores A, B, C e N' de Potvin *et al.* (1988), tensão *in-situ* estimada (σ_v , em MPa), ângulo formado entre a foliação e o plano do *hanging wall* (Alfa, em graus), índice de criticidade das estruturas geológicas singulares (ICEGs) e relações entre cabeamento de lavra e comprimento do realce (Mc/str), área do realce (Mc/Ahw) e tonelage do realce (Mc/ton);
- Parâmetros de perfuração e desmonte: perfuração específica (PE), proporção entre os furos espetando o *hanging wall* e o total de furos (FEhw) e proporção entre os furos espetando o *footwall* e o total de furos de perfuração (FEfw);
- Diluição do realce (Diluição, em %).

Uma descrição detalhada de cada parâmetro pode ser obtida em Rodrigues (2023).

3.2 Metodologia

Após tratativa do banco de dados, inicia-se a investigação de importância das variáveis. Árvores de decisão (DT) e Florestas Aleatórias (RF) foram os métodos de inteligência artificial escolhidos para tal, uma vez que são adequados na definição de importância de variáveis devido à estrutura hierárquica das DT e a capacidade analítica de influência das variáveis do estimador OOB nas RF. Todas as análises foram realizadas no R v. 4.1.2 (R CORE TEAM, 2021), por meio do ambiente integrado de desenvolvimento RStudio v. 2023.12.1 (RSTUDIO TEAM, 2023). Os seguintes pacotes foram utilizados: plyr (WICKHAM, 2011);

randomForest (LIAW; WIENER, 2002); rpart (THERNEAU; ATKINSON, 2019); rpart.plot (MILBORROW, 2021).

Tendo conhecimento das particularidades e diferenças geológico-geotécnicas entre as mineralizações Carvoaria (CV), Cachorro Bravo (CB) e Laranjeiras (LJ), propõe-se que sejam feitas análises para os três corpos de forma separada. Contudo, devido ao baixo número de amostras, o RStudio não gerou resultados para as análises por DT. Foi então proposto que todos os realces do banco de dados fossem avaliados em conjunto numa série de análises que alternaram sementes aleatórias pré-definidas, proporções treino/teste e parâmetros de *tunning*, num total de 200 análises computacionais, visando a observação de importância de variáveis em cenários completamente diferentes entre si. Em cada análise foi feita contagem e registro das aparições das variáveis utilizadas na construção das DT. Ao final, as seis mais recorrentes foram aplicadas nas análises seguintes. Este processo pode ser melhor compreendido na leitura de Rodrigues (2023), que também apresenta como os hiperparâmetros a serem utilizados nos modelos de RF foram definidos.

Em seguida são realizadas 50 análises para cada mineralização, alterando apenas as sementes pré-definidas, e registrados os valores das métricas de validação em cada análise. As métricas de validação selecionadas para validar os modelos preditivos são o coeficiente de determinação R^2 e a raiz do erro quadrático médio RMSE, e tiveram sua capacidade de generalização medida pelo método da ressubstituição com divisão amostral (*random subsampling*). O mesmo modelo de RF foi utilizado para definição de importância de variáveis a partir do estimador OOB. A forma de amostragem aleatória das RF possibilita que o estimador OOB seja utilizado mesmo que haja baixo número de amostras no banco de dados, utilizando todas as variáveis e amostras disponíveis no banco de dados.

A fim de comparar os dois métodos, foram registradas a importância de todas as variáveis em cada mineralização, e nos três casos, escolhidas as seis com maiores valores apontados pelo estimador OOB. Foram então feitas 50 novas análises para cada mineralização, com as mesmas sementes pré-definidas, agora utilizando as variáveis indicadas pelo estimador OOB em cada mineralização, e registradas as métricas de validação para que sejam comparados os métodos e escolhido aquele que consegue ter maior capacidade preditiva, com melhor acurácia, utilizando as variáveis mais importantes do banco de dados.

4 RESULTADOS E DISCUSSÕES

As avaliações por DT indicam que as variáveis que mais figuraram no total de análises foram espessura planejada (Ep), ângulo formado entre a foliação e o plano do *hanging wall* (Alfa), largura do *undercutting* no *hanging wall* da galeria inferior (LUGIhw), profundidade (Prof), índice de criticidade das estruturas geológicas singulares (ICEGs) e furos espetando o *footwall* (FEfw), em ordem de significância por quantidade de aparições. Chama a atenção a ocorrência de variáveis de todos os grupos de parâmetros, indicando que de fato são diversos os fatores que controlam a diluição, e uma análise aprofundada é requerida para contemplar todos esses fatores.

De posse das variáveis mais significativas, os modelos de RF são gerados para os bancos de dados de cada uma das mineralizações, e avaliadas suas métricas de validação e ordem de importância de variáveis. O modelo preditivo do corpo Carvoaria foi denominado RFDT-CV, do corpo Cachorro Bravo RFDT-CV, e do corpo Laranjeiras RFDT-LJ. A Tabela 1 apresenta os valores das métricas de validação obtidas, e a Tabela 2 apresenta a ordem de significância das variáveis pelas análises por DT e dos modelos RFDT-CV, RFDT-CB e RFDT-LJ.

Tabela 1. Resultados dos modelos de florestas aleatórias para banco de dados de cada corpo utilizando variáveis mais relevantes obtidas por DT.

MODELO	RMSE TREINO	RMSE TESTE	R ² TREINO	R ² TESTE
RFDT-CV	5.1156	11.7239	0.9231	0.5295
RFDT-CB	3.5481	7.8609	0.8846	0.5465
RFDT-LJ	4.8513	10.0041	0.9164	0.4525

Tabela 21. Comparação entre as importâncias das variáveis nos diferentes modelos de aprendizado de máquina para variáveis mais importantes obtidas por DT.

ÁRVORES DE DECISÃO		FLORESTAS ALEATÓRIAS		
		CARVOARIA	CACHORRO BRAVO	LARANJEIRAS
1	Ep	Ep	Ep	Alfa
2	Alfa	Alfa	Alfa	Ep
3	LUGIhw	LUGIhw	LUGIhw	ICEGs
4	Prof	Prof	ICEGs	FEfw
5	ICEGs	FEfw	Prof	Prof
6	FEfw	ICEGs	FEfw	LUGIhw

Utilizando o mesmo algoritmo de RF o estimador OOB foi aplicado para cada mineralização, utilizando agora todas as 26 variáveis do banco de dados tratado, a fim de observar possíveis semelhanças e discrepâncias entre os métodos na definição de importância de variáveis, e selecionadas as seis de maior significância para que fossem realizadas novas análises nos algoritmos de cada mineralização, a fim de obter as métricas de validação e comparar com aquelas obtidas anteriormente e apresentadas na Tabela 1. Os modelos agora são denominados RFOOB-CV para o corpo Carvoaria, RFOOB-CB para o corpo Cachorro Bravo e RFOOB-LJ para o corpo Laranjeiras. A Tabela 3 apresenta os valores das métricas de validação obtidas, enquanto a Tabela 4 apresenta a sequência de importância de variáveis para cada mineralização a partir dos modelos RFOOB.

Tabela 3. Resultados dos modelos de florestas aleatórias para banco de dados de cada corpo utilizando variáveis mais relevantes obtidas pelo estimador OOB.

MODELO	RMSE TREINO	RMSE TESTE	R ² TREINO	R ² TESTE
RFOOB-CV	3.3633	7.9299	0.8957	0.5519
RFOOB-CB	4.6610	10.6139	0.9293	0.6413
RFOOB-LJ	4.6725	9.6966	0.9156	0.5008

Tabela 42. Comparação entre as importâncias das variáveis nos diferentes modelos de aprendizado de máquina para variáveis mais importantes obtidas por OOB.

	CARVOARIA	CACHORRO BRAVO	LARANJEIRAS
1	Ep	Ep	Alfa
2	RMR	RMR	PE
3	Mc/ton	Alfa	Ep
4	LUGIhw	Prof	C
5	Q'	σ	HwDip
6	N'	A	GSI

A análise de importância de variáveis pelo estimador OOB indicou algumas semelhanças com aquela obtida pela DT, principalmente da significância das variáveis Ep e Alfa como as mais relevantes em correlação com diluição operacional, para os três casos. Contudo, é interessante observar as particularidades de cada mineralização quando utilizado o estimador OOB, possibilitando que problemas específicos das minas que operam nos diferentes corpos sejam avaliados e tratativas sejam direcionadas para eles. Além disso, observa-se que os valores de R² nos três casos foram superiores nos modelos RFOOB, o que indica que as variáveis indicadas como mais relevantes possuem maior capacidade descritiva nas correlações com a diluição quando comparado com os modelos RFDT, que tendem a descrever a diluição de modo mais generalizado uma vez que consideram todos os realces juntos para sua análise.

As variáveis Ep e Alfa constam como as mais relevantes em todas as análises, especialmente nas realizadas por DT, as duas mais relevantes em todos os casos. Considerando que a mina em estudo possui mineralizações caracterizadas por veios estreitos, espera-se que o parâmetro Ep seja um dos mais

significativos, pois pequenas variações da espessura planejada podem significar uma alta porcentagem de diluição. A relevância da variável Alfa é explicada pela presença de foliação nas mineralizações. Uma vez que o ângulo formado entre a foliação e o *hanging wall* dos realces é menor que 30° na maioria dos casos, observa-se uma tendência a maior instabilidade no realce pela ocorrência de deslocamentos após a abertura da escavação de lavra, principalmente devido ao alívio de tensões nessas regiões.

Observando de modo particular as variáveis mais relevantes obtidas pelos modelos RFOOB, seguem observações pertinentes de cada uma. No corpo Carvoaria é observada uma correlação direta entre as outras quatro variáveis mais relevantes na geração de diluição operacional. As classificações geomecânicas RMR e Q' possuem os menores valores médios neste corpo dentre os três. Os menores valores de Q' impactam diretamente no valor de N' , uma vez que é um dos quatro fatores utilizados na obtenção deste indicador. De acordo com o modelo empírico de Potvin *et al.* (1988), menores valores de N' devem ter valores de raio hidráulico correspondente também mais baixos para que haja estabilidade do realce. Contudo, os dados indicam que os valores de raio hidráulico praticados nos realces do corpo Carvoaria são os maiores dentre os três. A proporção de metros de cabeamento por tonelada de massa (Mc/ton) deste corpo é a maior entre as três, numa tentativa de trazer estabilidade aos realces de elevado valor de raio hidráulico, contudo, a variável também foi apontada como uma das mais relevantes na ocorrência de diluição. Assim, sugere-se que as geometrias dos realces praticados no corpo Carvoaria sejam mais conservadoras, com menores alturas ou extensão, a fim de minimizar o raio hidráulico, uma vez que as classificações geomecânicas apontam para menores características de resistência do maciço rochoso. O cabeamento de lavra deve ser acompanhado em campo para certificar que o procedimento está sendo bem executado, e a equipe técnica deve estudar possíveis modificações ou adequações no procedimento para que seja mais efetivo de acordo com as realidades da mina, como a mudança de malha de cabeamento, alteração nos tipos de cabo ou método de instalação/injeção.

A análise das variáveis mais relevantes do corpo Cachorro Bravo evidencia que um dos principais indicadores dos problemas de diluição dos realces executados neste corpo está associado às tensões, sejam elas *in-situ* ou induzidas. Inicialmente, observa-se que as variáveis Prof e σ_v ocorrem como mais relevantes e possuem correlação proporcional direta, posto que é esperado que as tensões *in-situ* aumentem com maiores profundidades. O fator A de Potvin, também uma das variáveis mais relevantes, indica a razão entre as tensões induzidas pela escavação e a resistência à compressão uniaxial da rocha intacta. Corpos foliados tendem a ter alívios de tensão nos planos de foliação, evento comumente observado no *hanging wall* de realces abertos. Desta forma, é importante que sejam conduzidos testes para conhecer profundamente como as tensões *in-situ* e induzidas se comportam no maciço rochoso do corpo Cachorro Bravo, visando a construção de banco de dados e confecção de modelagens geotécnicas confiáveis e robustas, que possam compreender valores e direções reais dos vetores de tensão no maciço rochoso. De posse desses dados, é imprescindível que haja um trabalho criterioso entre as equipes de geotecnia e planejamento de mina na definição do posicionamento e geometria dos realces, além de observações de campo para visualizar os efeitos das tensões nos realces.

As análises do corpo Laranjeiras chamam atenção pois, além da importância de variáveis diferir bastante daquelas dos outros dois corpos tanto nas análises de DT quanto nas OOB, os valores de R^2 são menores que das análises de Carvoaria e Cachorro Bravo. Isso pode indicar que as variáveis do banco de dados são menos descritivas na definição da diluição, havendo necessidade de adicionar novos parâmetros ao banco de dados, visando ampliar as correlações entre variáveis e também com a diluição operacional. Ainda assim, algumas observações podem ser feitas quanto aos resultados encontrados. Sendo o GSI uma classificação geomecânica que descreve o maciço quanto à estrutura visual da qualidade da superfície e formação de blocos, há evidência de grau de fraturamento médio a elevado neste corpo de acordo com os valores do banco de dados. Atrélado à isso está a aparição da variável C de Potvin entre as mais relevantes, indicando que o modo de falha por queda de blocos possui correlação forte na geração de *overbreak* deste corpo, especialmente na ocorrência de descontinuidades sub-horizontalizadas. Além da variável Alfa ser a mais relevante, tem-se que o ângulo de inclinação do *hanging wall* do realce também figura entre os fatores causadores de diluição nos realces do corpo. Por fim, a variável PE, definida como a razão entre a massa total planejada do realce pela metragem total do projeto de perfuração, figura entre as mais significantes. É necessário que haja dimensionamento do plano de fogo adequado às condições geoestruturais dos realces desse corpo, adotando estratégias que tragam menos perturbações ao maciço rochoso, principalmente para que as falhas por queda de blocos não sejam provocadas. É de suma importância que haja bom conhecimento das descontinuidades do corpo para este fim, com constantes observações em campo e atualização dos modelos geotécnicos.

Quanto às variáveis mais significativas pelas análises de DT, os resultados mostram que as escolhas operacionais para dimensionamento de galerias, erros durante a sua execução e a falta de aderência ao plano de perfuração e desmonte criam condicionantes que provocam diluição – a ocorrência de uma variável de largura de *undercutting* e outra da relação de furos de desmonte espetando o *footwall* indicam isso. O controle operacional destes fatores deve ocorrer visando a redução destas anomalias e consequente controle da diluição.

O alto valor da métrica RMSE teste em todos os casos indica que ainda há margem para desvios preditivos significativos nos modelos. Este inconveniente certamente pode ser contornado com a construção de um banco de dados com mais elementos amostrais de cada mineralização. É esperada uma maior capacidade preditiva na ocorrência de mais realces no banco de dados, pois o algoritmo conseguirá propor análises de correlação entre as variáveis e a diluição cada vez mais robustas e com menor erro – ou seja, maiores valores de R^2 teste. Contudo, apesar da necessidade de ampliar o banco de dados e da necessidade de aumentar os valores de acurácia do indicador R^2 e minimizar as variações do RMSE, o objetivo de avaliar a importância de variáveis consegue ser atingido, apontando para os principais desafios operacionais e geológico-geotécnicos que cada corpo propõe durante sua exploração.

5 CONCLUSÕES

A análise de importância de variáveis pelo estimador de erros OOB apresentou maiores valores de R^2 do que as análises por DT, sendo mais compatíveis com o atual banco de dados na interpretação das variáveis disponíveis em sua correlação com a diluição operacional. Dada a restrição na quantidade de amostras quando os corpos Carvoaria, Cachorro Bravo e Laranjeiras são individualizados em bancos de dados distintos, o estimador OOB se mostrou mais robusto na análise de significância de variáveis, visto que as DT só puderam realizar esta análise com utilização de todo o banco de dados.

Os dois tipos de análise indicam que as variáveis E_p e Alfa são as que possuem maiores correlações com a diluição, em todas as mineralizações. Cada uma apresenta suas particularidades quando observadas a importância das variáveis, apontando para a necessidade de diferentes abordagens na tratativa das questões de diluição em cada caso. Enquanto nos realces do corpo Carvoaria há um aparente mal dimensionamento da geometria das escavações, no corpo Cachorro Bravo as tensões *in-situ* e induzidas demonstram ser as principais condicionantes de diluição. Já no corpo Laranjeiras é necessário ter um controle operacional bastante acurado, visando aprimorar os planos de fogo, de modo que sua execução não provoque as rupturas por queda de blocos esperadas nesta mineralização.

Constata-se que os valores de RMSE teste são muito elevados, dando margem para grandes variações dos modelos preditivos, e a necessidade de aprimorar os valores de R^2 dos modelos, visando previsões mais acuradas para cada mineralização. Os valores de R^2 obtidos pelas análises RFOOB são de 0.5519 para o corpo Carvoaria, 0.6413 para o corpo Cachorro Bravo e de 0.5008 para o corpo Laranjeiras. Em ambos os casos se observa a necessidade de ampliar a quantidade de amostras nos bancos de dados.

O objetivo do trabalho, que consistiu na análise de variáveis para o estudo de caso em questão, foi alcançado. Além disso, este estudo abre portas para pesquisas futuras que busquem aprimorar a aplicabilidade do aprendizado de máquina na mineração subterrânea. Ampliar a amostragem, reduzir a complexidade dos modelos ou facilitar sua compreensão, explorar variáveis ocultas e suas interações, e investigar alternativas para reduzir o RMSE são apenas alguns exemplos de caminhos promissores a serem explorados.

AGRADECIMENTOS

Agradeço aos professores Dr. Allan Erlinkhman e MSc. Guilherme Alzamora por ajudar de tantas formas para que este trabalho pudesse ser finalizado. À UFOP, por proporcionar a oportunidade da execução desta pesquisa e também todos os recursos necessários.

REFERÊNCIAS BIBLIOGRÁFICAS

- Breiman, L. (2001) Random forests. *Machine Learning*, vol 45(1), p. 5–32.
- Charbel, P. A. (2015) *Gerenciamento de Risco Aplicado à Diluição de Minério*. Tese de doutorado. Publicação / Departamento de Engenharia Civil e Ambiental, Universidade de Brasília, Brasília, DF, 406 p.
- Clark, L. M. (1998) *Minimizing dilution in open stope mining with a focus on stope design and narrow vein longhole blasting*. University of British Columbia, Vancouver, Canadá.

- Costa, L. C. B. (2017) *Metodologia para prever e reduzir o risco de diluição não planejada em lavras estreitas pelo método sublevel stoping. Estudo de caso: mina Córrego do Sítio, Santa Bárbara/MG*. Dissertação de mestrado. Mestrado Profissional em Geotecnia, Universidade Federal de Ouro Preto.
- Guedes, E. V. O. (2020) *Aplicação de soft sensor baseado em redes neurais artificiais e random forest para predição em tempo real do teor de ferro no concentrado da flotação de minério de ferro*. Dissertação de mestrado. Programa de Pós-Graduação em Instrumentação, Controle e Automação de Processos de Mineração, Universidade Federal de Ouro Preto e Associação Instituto Tecnológico Vale. Ouro Preto.
- Jang, H.; Topal, E.; Kawamura, Y. (2015) Unplanned dilution and ore loss prediction in longhole stoping mines via multiple regression and artificial neural network analyses. *The Journal of The Southern African Institute of Mining and Metallurgy*, volume 115, p. 449-456.
- Lemos, E. P.; Arns, M. T. S.; Nievola, J. C. (2005) Análise de crédito bancário por meio de redes neurais e árvores de decisão: uma aplicação simples de data mining. *Revista de Administração: RAUSP*, vol. 40, núm. 3, julio-septiembre, pp. 225-234. Universidade de São Paulo, São Paulo, Brasil.
- Liaw, A.; Wiener, M. (2002) Classification and Regression by randomForest. *R News* 2(3), 18--22.
- Milborrow, S. (2021) rpart.plot: Plot 'rpart' Models: An Enhanced Version of 'plot.rpart'. R package version 3.1.0. Disponível em: <<https://CRAN.R-project.org/package=rpart.plot>>. Acesso em fev. 2023.
- Oshiro, T. M. (2013) *Uma abordagem para a construção de uma única árvore a partir de uma Random Forest para classificação de bases de expressão gênica*. Dissertação (Mestrado em Bioinformática): Bioinformática, Universidade de São Paulo, São Paulo.
- Potvin, Y.; Hudyma, M. (1988) *The stability graph method for open stope design*. Paper nº44, University of British Columbia, Edmonton, Alberta, Canada,
- Qi, C.; Fourie, A.; Du, X.; Tang, X. (2018) Prediction of open stope hangingwall stability using random forests. Artigo. *Natural Hazards*, volume 92, p. 1179-1197.
- R Core Team (2021) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <<https://www.R-project.org/>>. Acesso em fev. 2023.
- Rodrigues, C. O. (2023) *Previsão da diluição em realces subterrâneos por meio de técnicas de aprendizado de máquina*. Dissertação (Mestrado em Engenharia Mineral) - Escola de Minas, Universidade Federal de Ouro Preto, Ouro Preto. 131 p.
- Rstudio Team (2023) RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA. Disponível em: <<http://www.rstudio.com/>>. Acesso em mar. 2024.
- Santos, A. E. M. (2021) *Classificação de maciços rochosos por meio de técnicas da estatística multivariada e inteligência artificial*. Tese de doutorado. Doutorado em Engenharia Mineral, Universidade Federal de Ouro Preto.
- Santos, T. B. (2019) *Análise de risco geotécnico em taludes rochosos de mina com uso de técnicas estatísticas multivariadas e de aprendizado de máquina*. Tese de doutorado. Doutorado em Engenharia Mineral, Universidade Federal de Ouro Preto.
- Therneau, T; Atkinson, B. (2019) rpart: Recursive Partitioning and Regression Trees. R package version 4.1-15. Disponível em: <<https://CRAN.R-project.org/package=rpart>>. Acesso em fev. 2023.
- Wickham, H. (2011) The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software*, 40(1), 1-29. Disponível em: <<http://www.jstatsoft.org/v40/i01/>>. Acesso em fev. 2023.