

DOI: 10.47094/COBRAMSEG2024/115

REVISÃO DE LITERATURA: UTILIZAÇÃO DE MACHINE LEARNING E DADOS GEOTÉCNICOS PARA PROJETOS DE DISPOSIÇÃO DE ESTERIL E REJEITO PARA FEL-1

Renata Delicio Andrade de Freitas
Engenheira Geotécnica, Progen, Nova Lima, Brasil, c0663568@vale.com

Bráulio Araujo Rodrigues
Engenheiro Geotécnico, Vale, Nova Lima, Brasil, braulio.rodrigues@vale.com

Jhonimar Lamarque Santos
Técnico em mineração, Progen, Nova Lima, Brasil, c0632814@vale.com

Luiz Roberto Schuchter Júnior
Engenheiro, Progen, Nova Lima, Brasil, c0623885@vale.com

Débora Louyse Alpes Melo
Engenheira Geotécnica, Progen, Nova Lima, Brasil, c0679781@vale.com

RESUMO: O FEL 1, também conhecido como *Front End Loading 1*, é uma etapa crucial no processo de desenvolvimento de projetos, especialmente no setor de mineração. Nessa fase, é realizada a identificação e validação da Oportunidade de Negócio, além de estudos e análises fundamentais para aprovação do projeto. No caso específico, o FEL 1 é dividido em quatro etapas, com o objetivo de selecionar a melhor área para implantação da estrutura. A primeira etapa consiste no levantamento de áreas potenciais com baixo DMT, considerando as características geomorfológicas do terreno. A segunda etapa envolve levantamento dos aspectos legais que possam impedir a alocação da estrutura. Na terceira etapa, são considerados os aspectos relacionados à licença social e aos compromissos da empresa. É essencial analisar o impacto do projeto na comunidade local e garantir que todas as exigências sejam atendidas. Por fim, na quarta etapa, ocorre o ranqueamento das alternativas. Nesse processo, é feita uma avaliação criteriosa das áreas identificadas nas etapas anteriores, considerando critérios como viabilidade técnica, econômica e ambiental. No entanto, foi constatado pelo grupo de "Pesquisa e Desenvolvimento" que o FEL 1 atualmente realiza a identificação dessas áreas de forma manual e sem utilizar dados geotécnicos para a seleção. Nesse contexto, o objetivo deste estudo é conduzir uma revisão da literatura com o propósito de estabelecer uma base sólida para a criação de um mapa de favorabilidade geotécnica. Esse mapa será utilizado pelo FEL 1 para classificar as áreas alternativas, empregando técnicas de *Machine Learning*. Serão revisados diferentes modelos, como *Random Forest*, Rede Neural, *Naive Bayes* e *Gradient Boosting*, a fim de identificar o algoritmo com melhor desempenho. Essa revisão servirá de ponto de partida para posterior realização de criação de um modelo de aprendizado de máquina que esperamos treinar com diversas fontes, como sensoriamento remoto, geofísica, geologia e geotecnia. É importante ressaltar que os parâmetros geotécnicos serão validados pelo programa RCA (Rastreabilidade, Confiabilidade e Armazenamento), que garante a qualidade das informações geotécnicas obtidas em campo. Com a implementação desse modelo de seleção de áreas baseado em dados geotécnicos confiáveis, bem como os demais dados, espera-se otimizar o processo de identificação e seleção da melhor área para implantação da estrutura de disposição durante a etapa do FEL 1. Isso resultará em uma tomada de decisão mais precisa e embasada em informações técnicas confiáveis, contribuindo para a eficiência, segurança geotécnica e sucesso do projeto.

PALAVRAS-CHAVE: FEL1, Machine Learning, Favorabilidade Geotécnica, RCA.

ABSTRACT: The FEL 1, also known as Front End Loading 1, is a crucial stage in the project development process, especially in the mining sector. During this phase, business opportunities are identified and validated, along with fundamental studies and analyses for project approval. In the specific case, FEL 1 is divided into four stages, with the goal of selecting the best area for implementation. The first stage involves surveying potential areas with low DMT (Dynamic Cone Penetration Test) based on the geomorphological characteristics of the terrain. The second stage includes legal assessments to ensure no impediments to structure allocation. In the third stage, social license aspects and company commitments are considered. It's essential to analyze the project's impact on the local community and ensure compliance with all requirements. Finally, in the fourth stage, alternatives are ranked. This process involves a thorough evaluation of the areas identified in the previous stages, considering technical, economic, and environmental viability criteria. However, the 'Research and Development' group has found that the current manual identification of these areas by FEL 1 does not utilize geotechnical data. In this context, the objective of this study is to conduct a literature review to establish a solid foundation for creating a geotechnical favorability map. This map will be used by FEL 1 to classify alternative areas using machine learning techniques. Different models, such as Random Forest, Neural Networks, Naive Bayes, and Gradient Boosting, will be reviewed to identify the algorithm with the best performance. This review will serve as a starting point for creating a machine learning model that we hope to train using various sources, including remote sensing, geophysics, geology, and geotechnics. It's important to note that geotechnical parameters will be validated by the TRS program (Traceability, Reliability, and Storage), ensuring the quality of geotechnical information obtained in the field. Implementing this area selection model based on reliable geotechnical data, along with other information, aims to optimize the identification and selection process for the structure's placement during the FEL 1 stage. This will result in more precise decision-making based on reliable technical information, contributing to efficiency, geotechnical safety, and project success.

KEYWORDS: FEL1, Machine Learning, Geotechnical Favorability, RCA.

1 INTRODUÇÃO

O *Front End Loading 1* (FEL 1) é responsável pela identificação e validação de áreas potenciais para a disposição de rejeito e estéril em pilhas. Este é um componente crítico do processo de desenvolvimento de projetos de mineração. Para garantir a segurança e eficiência das estruturas de disposição, bem como os requisitos geotécnicos e ambientais, essas áreas devem ser escolhidas com cuidado. A identificação de áreas favoráveis para uma análise abrangente e precisa neste contexto requer o uso de técnicas de *Machine Learning* (ML) sofisticadas combinadas com dados geotécnicos. A combinação dessas ferramentas tem o objetivo de revolucionar a abordagem tradicional do FEL 1, permitindo uma tomada de decisão embasada em informações técnicas confiáveis e aumentando significativamente o sucesso e a segurança dos projetos.

2 METODOLOGIA DE MACHINE LEARNING APLICADO À SELEÇÃO DE ÁREAS DE DISPOSIÇÃO DE ESTEREIS E REJEITO

Em vários campos, como engenharia, ciência de dados e geotecnia, o uso de técnicas de *Machine Learning* (ML) tem se mostrado uma abordagem útil para resolver uma variedade de problemas complexos. Aprendizado de máquina é um conjunto de técnicas e algoritmos que permitem que sistemas computacionais façam previsões, tomem decisões e aprendam padrões de dados sem serem explicitamente programados para tais tarefas (Mitchell, 1997). Esses algoritmos podem ser treinados com conjuntos de dados para reconhecer padrões e relacionamentos, o que os torna particularmente úteis em problemas de classificação, onde o objetivo é atribuir uma categoria ou rótulo a uma entrada específica de acordo com suas características.

O uso de ML na seleção de áreas para projetos de disposição de estéreis e rejeito de mineração, como FEL 1, é pioneiro e importante porque esses algoritmos têm a capacidade de processar grandes quantidades de dados e reconhecer padrões complexos que podem ser difíceis de identificar por métodos tradicionais. As variáveis ambientais e socioeconômicas já são amplamente utilizadas como filtro de seleção para essas áreas, mas este trabalho tem como objetivo incluir também variáveis geotécnicas, que são essenciais e não devem ser negligenciadas.

Considerando que a fase FEL 1 não tem escopo de pesquisa direta de coletas de dados de campo, a utilização de *Machine Learning* s apresenta como ferramenta indispensável para prever parâmetros diretos, como NSPT, utilizando dados indiretos de geofísica e sensoriamento remoto. ML permite a análise simultânea de vários fatores e a criação de modelos preditivos, que podem ajudar na determinação das regiões mais favoráveis para a implantação dessas estruturas.

Estudos anteriores têm mostrado que as técnicas de *Machine Learning* podem ser úteis na seleção de áreas para uma variedade de propósitos, como mineração, agricultura, conservação ambiental e planejamento urbano. Por exemplo, Wu *et al.* (2017) descobriram áreas potenciais para mineração de ouro usando dados geológicos e ambientais. Eles usaram algoritmos de aprendizado de máquina como Support Vector Machine (SVM) e Random Forest. Os resultados mostraram que os modelos de aprendizado de máquina superaram os métodos convencionais de avaliação de áreas, fornecendo uma maior precisão na localização de regiões promissoras para exploração mineral.

Outro exemplo é o trabalho de Smith *et al.* (2019), os autores realizaram um estudo relevante que usa técnicas de ML para mapear precisamente áreas propícias à agricultura usando dados geoespaciais e ambientais. Os autores classificaram vários tipos de solos e descobriram locais com maior potencial agrícola usando algoritmos de aprendizado de máquina como Redes Neurais Artificiais e Boosting Gradiente. Os resultados mostraram que esses modelos são muito bons para demarcar zonas de manejo agrônomico e ajudar a otimizar o uso de recursos e aumentar a produtividade agrícola.

Esses exemplos são alguns dos que embasam este estudo. É entendida ampla aplicabilidade e os benefícios potenciais do uso de técnicas de *Machine Learning* na seleção de áreas para projetos de disposição de rejeitos e estéreis; eles fornecem insights valiosos para a tomada de decisão e ajudam a dar robustez ao resultado de FEL 1.

3 REVISÃO DE MÉTODOS DE MACHINE LEARNING

Este estudo examinará quatro dos algoritmos mais populares encontrados na literatura para resolver problemas de classificação que os métodos de aprendizado de máquina oferecem: Random Forest, Rede Neural, Naive Bayes e Gradient Boosting. Cada um desses métodos tem suas próprias características, que podem decidir se é útil ou eficaz na escolha de áreas para o projeto de disposição de materiais.

3.1 Random Forest

O Random Forest é um algoritmo de aprendizado de máquina baseado em árvores de decisão. Durante o treinamento, ele cria uma coleção de árvores de decisão e, em seguida, combina suas previsões para obter uma predição final (Breiman, 2001). Uma das principais vantagens do Random Forest é sua capacidade de lidar com conjuntos de dados grandes e complexos, bem como sua menor probabilidade de *overfitting* em comparação com árvores de decisão individuais.

Uma grande vantagem do algoritmo Random Forest é a sua capacidade de lidar com conjuntos de dados grandes e complexos, tornando-o uma opção confiável para uma variedade de problemas reais. Além disso, seu ponto forte é sua capacidade de fornecer resultados mais confiáveis e duradouros, reduzindo a probabilidade de *overfitting*.

Mas o Random Forest tem uma desvantagem: o modelo que ele produz é difícil de entender. Ao contrário de modelos mais simples, a combinação de várias árvores de decisão pode dificultar a compreensão do processo de tomada de decisões. Além disso, em comparação com modelos de árvore de decisão individuais, treinar um elevado número de árvores pode exigir mais tempo e recursos computacionais, embora na maioria das vezes seja eficaz. Portanto, ao usar o Random Forest, é fundamental levar em consideração esses elementos e avaliar a relação entre desempenho e interpretabilidade, conforme necessário para cada uso particular.

3.2 Rede Neural

As redes neurais artificiais são modelos computacionais que imitam como o cérebro humano funciona. Elas são compostas por várias camadas de neurônios interconectadas, cada uma com suas propriedades de ativação e ponderação. Devido à sua flexibilidade e capacidade de lidar com grandes quantidades de dados, as redes neurais são frequentemente usadas em problemas de classificação. Isso se deve à sua capacidade de aprender padrões complexos nos dados (Goodfellow *et al.*, 2016).

Uma grande vantagem das redes neurais artificiais é sua capacidade de aprender e modelar padrões complexos em grandes conjuntos de dados. Essa capacidade torna as redes neurais artificiais especialmente eficazes em problemas de classificação envolvendo grandes conjuntos de dados. Em comparação com métodos mais tradicionais, sua flexibilidade é uma grande vantagem, pois permite a adaptação a uma ampla variedade de tarefas. Além disso, é capaz de lidar com dados de alta dimensionalidade e não-lineares.

Uma desvantagem das redes neurais artificiais, no entanto, é a complexidade delas e a necessidade de processar uma grande quantidade de dados para treinamento. Além disso, ajustar os parâmetros da rede neural pode ser difícil e requer experiência técnica especializada. A *overfitting* das redes neurais também pode ocorrer em certas situações, especialmente quando os dados de treinamento são insuficientes. Portanto, a aplicação de redes neurais artificiais pode exigir uma quantidade significativa de recursos computacionais, apesar de sua potência.

3.3 Naive Bayes

O teorema de Bayes e a suposição "ingênua" de independência condicional entre os recursos são a base do algoritmo de aprendizado supervisionado *Naive Bayes* (Rish, 2001). O Naive Bayes é conhecido por sua eficiência computacional e é frequentemente usado em problemas de classificação com conjuntos de dados de alta dimensionalidade, embora essa suposição simplificadora possa não ser verdadeira em muitas situações.

Uma grande vantagem do algoritmo *Naive Bayes* é sua eficiência computacional, o que o torna adequado para trabalhar com conjuntos de dados com grande dimensão. Além disso, o Naive Bayes pode produzir resultados satisfatórios em vários casos práticos, especialmente em questões de classificação, mesmo com a suposição "ingênua" de independência condicional entre os recursos.

No entanto, a principal desvantagem do *Naive Bayes* está na suposição de independência condicional, que pode não ser verdadeira em muitas situações. Em algumas situações, a perda de precisão pode ser causada por essa simplificação, especialmente quando os recursos do conjunto de dados estão interconectados. Portanto, apesar da sua eficiência computacional e resultados satisfatórios em muitos casos, é importante considerar cuidadosamente a validade da suposição "ingênua" ao aplicar o algoritmo *Naive Bayes* em problemas reais.

3.4 Gradient Boosting

O *Gradient Boosting* é um método de aprendizado em grupo que cria um modelo de previsão a partir de um conjunto de modelos de aprendizado fracos, normalmente árvores de decisão rasas (Friedman, 2001). Ele funciona usando treinamento sequencial de cada modelo. Isso significa que cada novo modelo é treinado para corrigir os erros cometidos pelo modelo anterior. O *Gradient Boosting* é famoso por lidar com conjuntos de dados desbalanceados e lidar com outliers.

Uma grande vantagem do método *Gradient Boosting* é a capacidade de lidar com conjuntos de dados inconsistentes e exceções. Isso o torna uma opção eficaz para problemas do mundo real em que a modelagem de dados pode ser difícil e complexa. Além disso, o treinamento contínuo de todos os modelos, que corrige os erros cometidos pelos modelos anteriores, geralmente resulta em um desempenho preditivo forte e preciso.

Mas a sensibilidade a hiperparâmetros e a possibilidade de *overfitting* são uma desvantagem do *Gradient Boosting*. Para obter o melhor desempenho do modelo possível, os hiperparâmetros devem ser configurados corretamente, o que pode exigir experimentação e ajustes minuciosos. Além disso, em comparação com outros métodos, o *Gradient Boosting* pode ser mais propenso a *overfitting* devido ao treinamento sequencial. Portanto, para evitar problemas de *overfitting* durante o treinamento, a complexidade do modelo deve ser monitorada e ajustada adequadamente.

3.5 Aplicação para a escolha de espaços para projetos de disposição

Esses métodos de aprendizado de máquina podem ser usados na seleção de áreas para projetos de disposição de rejeito e estéreis. Conjuntos de dados sobre as áreas em estudo, incluindo informações geotécnicas, ambientais e socioeconômicas, serão usados para prever a favorabilidade de cada local para a instalação de pilha. Além disso, eles serão combinados com técnicas de processamento de imagens de sensoriamento remoto e métodos geofísicos já disponíveis para melhorar a precisão das previsões e identificar padrões espaciais complexos que podem afetar a escolha das áreas mais adequadas.

4 DADOS INDIRETOS E DIRETOS E SUA IMPORTÂNCIA NA SELEÇÃO DE ÁREAS PARA DISPOSIÇÃO DE REJEITOS E ESTÉREIS EM PILHAS DURANTE O FEL 1

Já sabemos da importância de FEL 1 como etapa de um projeto de disposição de rejeitos em pilha. Para esse estudo, os modelos de ML serão treinados utilizando dados indiretos, de sensoriamento remoto e geofísica aeroespacial, e dados diretos de campo já disponíveis. Os dados diretos servirão de base para o treinamento do modelo que será base, a partir da escolha da melhor metodologia, para produzir um mapa de calor que mostra com certa confiabilidade regiões mais favoráveis à implantação de uma estrutura dessas.

4.1 Sensoriamento Remoto

Por meio da captura de dados de superfície e imagens, o sensoriamento remoto permite a identificação de características geológicas, geomorfológicas e ambientais relevantes para a seleção de áreas. Por exemplo, imagens de satélite podem fornecer informações sobre o relevo, a drenagem, a cobertura vegetal e outros elementos do terreno que podem afetar a decisão de localizar estruturas em pilhas de rejeitos e estéreis.

Além disso, a análise de imagens de sensoriamento remoto permite a identificação de áreas que podem apresentar instabilidade geotécnica, como falhas, inclinações acentuadas e áreas suscetíveis a processos erosivos. Para avaliar os riscos geotécnicos associados à locação de estruturas em pilhas de rejeitos e estéreis, essas informações são essenciais. Elas também ajudarão a minimizar os impactos ambientais e a garantir a segurança das operações mineradoras.

Devido à sua capacidade de fornecer dados precisos e atualizados sobre as características do terreno, o sensoriamento remoto tem sido amplamente utilizado na indústria de mineração. O uso do sensoriamento remoto é crucial para avaliar os impactos ambientais associados à mineração e identificar áreas propícias à mineração (Srivastava et al., 2019; Lu et al., 2020).

4.2 Geofísica

Ao fornecer informações não invasivas sobre as características do subsolo, a geofísica aeroespacial complementa o sensoriamento remoto. Usando métodos como magnetometria, gravimetria e espectrometria de raios gama, é possível mapear a distribuição dos materiais geológicos, identificar estruturas geológicas e delinear unidades litológicas em profundidade.

A seleção de áreas para a instalação de estruturas em pilhas de rejeitos e estéreis depende especialmente do mapeamento geofísico do subsolo, pois permite avaliar a composição geológica e a estabilidade do terreno em profundidade. A geofísica aeroespacial também pode ser usada para descobrir se há recursos hídricos subterrâneos, como lençóis freáticos e aquíferos, que podem afetar a gestão de águas e a viabilidade das operações mineradoras.

Estudos recentes mostram que a geofísica aeroespacial é importante para a caracterização do subsolo e para a determinação de locais adequados para mineração. Estudos de Li et al. (2020) e Smith et al. (2021) mostraram que as técnicas geofísicas aeroespaciais são úteis para identificar depósitos minerais e avaliar a estabilidade geotécnica do terreno.

4.3 Relevância dos Dados Geotécnicos no FEL 1

Como fornecem informações importantes sobre as características do terreno e sua adequação para a disposição de rejeitos e estéreis em pilhas, os dados geotécnicos são essenciais para orientar a seleção de áreas durante o FEL 1. A capacidade de suporte do solo, a estabilidade do terreno e outros fatores geotécnicos que afetam diretamente a segurança e a eficiência das estruturas podem ser avaliados com esses dados.

Ao longo do FEL 1, a análise dos dados geotécnicos é essencial para identificar áreas com condições geotécnicas adequadas para a construção de pilhas de rejeitos e estéreis. Ao garantir a segurança das estruturas e do ambiente circundante, essa análise pode ajudar a reduzir riscos geotécnicos como instabilidade de taludes, subsidência e erosão.

4.4 Tipos de Dados Geotécnicos Aplicados para favorabilidade de FEL 1

Para começar, foram identificados três parâmetros geotécnicos essenciais para relacionar com os dados indiretos obtidos após consultas cuidadosas com especialistas em geotecnia. O NSPT e as informações do nível de água sobre o topo rochoso estão entre esses parâmetros. Esses parâmetros foram escolhidos porque são importantes para avaliar a competência do solo a permeabilidade do solo e outras características geotécnicas relevantes para escolher áreas adequadas para a disposição de rejeito e estéril em pilhas.

Além disso, é importante destacar que a lista de parâmetros pode ser ampliada caso mais informações sejam adicionadas para aumentar a precisão e a previsibilidade do modelo geotécnico.

5 RESULTADOS ESPERADOS E CONCLUSÃO

Com base na exaustiva revisão da literatura realizada, almejamos que este estudo se torne uma ferramenta valiosa para inferir, com o máximo de assertividade possível, os parâmetros geotécnicos em uma área ainda de *greenfield* ou com pouca investigação geotécnica. Esperamos que por fim seja possível que a resposta do modelo de ML produza um mapa de calor apostando áreas mais favoráveis à implementação de estruturas de pilha de rejeito e estéril. Vale lembrar que esse será um dos parâmetros utilizados na tomada de decisão de FEL 1 que já utiliza parâmetros ambientais e sociais como dados para negatizar áreas.

Por fim, acredita-se que a utilização dessas ferramentas pode permitir uma análise completa e precisa de áreas potenciais, fornecendo informações importantes para a tomada de decisão. O uso dos dados diretos e indiretos como inteligência geotécnica também abre caminho para uma discussão importante que é a qualidade e usabilidade dos dados coletados hoje. Por isso o trabalho visa utilizar dados geotécnicos validados pelo programa RCA para garantir a confiabilidade do resultado.

REFERÊNCIAS BIBLIOGRÁFICAS

- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189-1232.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.
- Li, X., Xu, J., Ma, S., Gu, J., He, Y. (2020). Application of Aero Geophysical Data in Prospecting and Geological Investigation in East and South China. *Geoscience*, 10(11), 1-12.
- Lu, H., Lu, Y., Zhu, J., Wu, J., Hu, J. (2020). Application of Remote Sensing Technology in the Monitoring of Ecological Environment in Mining Area. *Journal of Geoscience and Environment Protection*, 8(9), 157-170.
- Mitchell, T. M. (1997). *Machine Learning*. McGraw Hill.
- Rish, I. (2001). An empirical study of the naive Bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence* (Vol. 3, No. 22, pp. 41-46).
- Smith, A., Brown, D., Johnson, E., Wilson, P. (2021). The Application of Geophysical Methods in the Search for Mineral Deposits. *Mining Engineering*, 73(3), 52-57.
- Smith, J., et al. (2019). Application of machine learning techniques to precision agriculture: a review. *IEEE Access*, 7, 1525-1536.

- Srivastava, P.K., Srivastava, R., Gupta, R.P., Gupta, M. (2019). Remote sensing and GIS-based approach for the assessment of mining-induced land subsidence and its impact on land-use/land-cover (LULC): a case study of Singrauli coalfields, India. *Environmental Earth Sciences*, 78(12), 349.
- Wu, H., et al. (2017). A review of spatial prediction methods and modeling techniques for regional gold resource assessment based on gold exploration data. *Journal of Geochemical Exploration*, 176, 139-154