



MINISTÉRIO DA EDUCAÇÃO  
Universidade Federal de Ouro Preto  
Escola de Minas – Departamento de Engenharia de Minas  
Curso de Graduação em Engenharia de Minas

---

**JOSÉ MATHEUS VIEIRA MATOS**

**MODO DE FALHA CRÍVEL EM BARRAGENS DE REJEITO DE  
MINERAÇÃO: UMA ANÁLISE DAS VARIÁVEIS CONDICIONANTES E  
PROPOSIÇÃO DE UM MODELO PREDITOR**

Ouro Preto

2023

MODO DE FALHA CRÍVEL EM BARRAGENS DE REJEITO DE MINERAÇÃO: UMA  
ANÁLISE DAS VARIÁVEIS CONDICIONANTES E PROPOSIÇÃO DE UM MODELO  
PREDITOR

José Matheus Vieira Matos

Monografia de conclusão de curso para  
obtenção do grau de Engenheiro de Minas  
na Universidade Federal de Ouro Preto  
defendida e aprovada em 23 de agosto de  
2023 como parte dos requisitos para a  
obtenção do Grau de bacharel em  
Engenharia de Minas.

Banca examinadora:

Dr<sup>a</sup> Tatiana Barreto dos Santos (orientadora) – UFOP

Dr. Allan Erlikhman Medeiros Santos – UFOP

Msc. Danielle Aparecida de Menezes – BHP Billiton

Área de concentração: Lavra de Minas

Ouro Preto

2023

## SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

M433m Matos, José Matheus Vieira.

Modo de falha crível em barragens de rejeito de mineração  
[manuscrito]: uma análise das variáveis condicionantes e proposição de  
um modelo preditor. / José Matheus Vieira Matos. - 2023.  
85 f.

Orientadora: Profa. Dra. Tatiana Barreto dos Santos.  
Monografia (Bacharelado). Universidade Federal de Ouro Preto.  
Escola de Minas. Graduação em Engenharia de Minas .

1. Lavra de minas. 2. Geotecnia. 3. Aprendizado de máquina. 4.  
Barragens de rejeito. I. Santos, Tatiana Barreto dos. II. Universidade  
Federal de Ouro Preto. III. Título.

CDU 624.136:004.85

Bibliotecário(a) Responsável: Sione Galvão Rodrigues - CRB6 / 2526



## FOLHA DE APROVAÇÃO

**José Matheus Vieira Matos**

**Modo de falha crível em barragens de rejeito de mineração: uma análise das variáveis condicionantes e proposição de um modelo preditor**

Monografia apresentada ao Curso de Engenharia de Minas da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de bacharel em Engenharia de Minas

Aprovada em 23 de agosto de 2023

### Membros da banca

Dr<sup>ª</sup>. Tatiana Barreto dos Santos - Orientador(a) - Universidade Federal de Ouro Preto  
Dr. Allan Erlichman Medeiros Santos - Universidade Federal de Ouro Preto  
Msc. Danielle Aparecida de Menezes - BHP Billiton

Tatiana Barreto dos Santos, orientadora do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 23/08/2023



Documento assinado eletronicamente por **Tatiana Barreto dos Santos, PROFESSOR DE MAGISTERIO SUPERIOR**, em 28/08/2023, às 13:55, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [http://sei.ufop.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **0581913** e o código CRC **8E7DC837**.

*Este trabalho é todo dedicado aos meus pais, pois é graças ao seu esforço que hoje posso concluir o meu curso.*

## **AGRADECIMENTOS**

Agradeço primeiramente a Deus por me conceder saúde, capacidade, paciência e sabedoria para realizar o sonho que a graduação sempre representou para mim.

Aos meus pais, Josiane e Leonardo, por terem sido o meu alicerce e terem me auxiliado não apenas durante o período de graduação, mas sim por todo meu caminho até atingi-la. Agradeço também, por todos os conselhos e momentos felizes que tivemos juntos, à Talita, Days, Rodrigo, Ianka, Nathália e Nicolay, amigos mais antigos que fiz nessa jornada na UFOP, além dos demais que acompanharam o meu crescimento e fizeram desta caminhada um trajeto mais leve e feliz, em especial Emanuel, Vitória, Duda, Lucas, Victor, Gabriela e Pedro.

Aos professores Ivair, Bruna Postacchini e Tatiana Barreto pelo apoio e auxílio no desenvolvimento de projetos de pesquisa. Também ao Vinícius Almeida, Kléber e Carlos Alberto, pelo vasto conhecimento que me transmitiram tanto em aulas como em orientações de projetos, e por contribuírem de maneira efetiva para o profissional que estou me tornando.

À Vale pela oportunidade de crescimento profissional, técnico e humano, especialmente à Malu Cecchetto pelo apoio e amizade.

Agradeço à Universidade Federal de Ouro Preto e à Escola de Minas, por todas as oportunidades de projetos, cursos e experiências engrandecedoras que tive em meu tempo como graduando. E à FAPEMIG, pelo auxílio no desenvolvimento do projeto de pesquisa APQ501-21, que possibilitou a elaboração desse trabalho.

## RESUMO

A importância de se conhecer o modo de falha crível em barragens de rejeito se dá à medida que tais falhas causam danos ambientais e econômicos consideráveis, além de constituírem fontes críticas de perdas de vidas humanas. Em vista disso, o conhecimento do modo de falha crível para uma dada estrutura é importante no seu gerenciamento de riscos, uma vez que, a depender do modo de falha, a análise de estabilidade geotécnica será diferente. Dentre os diferentes modos de ruptura, tem-se: instabilidade de taludes, erosão interna, liquefação, galgamento, inadequações estruturais ou na fundação, abalo sísmico e erosão externa. Neste trabalho foi realizado o estudo de algumas das variáveis relacionadas ao fenômeno e foram elaborados modelos preditores para modos de falha. Para tanto, fez-se uso da técnica de aprendizado de máquina *K Nearest Neighbors*, via linguagem *Python*. O desenvolvimento do estudo se deu a partir de um banco de dados com 66 barragens de rejeito rompidas do mundo todo. As variáveis conhecidas para essas barragens incluem o tipo de minério que gerou o rejeito armazenado, o material de construção da barragem, o método construtivo, altura, volume, risco sísmico, clima e modo de falha. Os resultados mostraram que todas as variáveis foram importantes na predição do modo de falha, exceto a variável "volume". Essa variável não demonstrou forte influência nos padrões de falha das barragens estudadas. Foram desenvolvidos trinta e cinco modelos preditores de modo de falha de barragens, identificados como M1 a M35. O modelo mais preciso foi o M11, no qual obteve-se uma acurácia de 71%. Esse modelo excluiu a variável de volume e foi configurado com k vizinhos igual a 3, tamanho proporção da amostra de treino/teste igual a 90%/10%. Tal modelo foi aplicado em um banco de dados com 10 barragens brasileiras classificadas pela Agência Nacional de Mineração com nível de emergência II e III. Dentre elas, oito foram classificadas para modo de falha crível por instabilidade de taludes e duas por erosão interna.

**Palavras-chave:** Risco geotécnico, Aprendizado de máquina, K vizinhos mais próximos, barragens de rejeito.

## ABSTRACT

The significance of understanding credible failure modes in tailings dams stems from the significant environmental and economic devastation such failures can cause. Moreover, these failures represent critical threats to human life. Consequently, grasping the credible failure mode of a particular structure holds importance in its risk management. Depending on the specific failure mode, the geotechnical stability analysis varies. Several failure modes exist, encompassing slope instability, internal erosion, liquefaction, overtopping, structural or foundation inadequacies, earthquakes, and external erosion. In this study, an examination of variables related to this phenomenon was conducted, and predictive models for failure modes were formulated. To achieve this, the K Nearest Neighbors machine learning technique was employed using the Python language. The study's foundation relied on a database featuring 66 instances of failed tailings dams across the globe. Notable variables encompassed the ore type generating the stored tailings, dam construction materials, construction methods, height, volume, seismic risk, climate, and failure mode. The findings demonstrated the significance of all variables in predicting failure modes, with the exception of the "volume" variable, which exhibited minimal influence on failure patterns in the analyzed dams. A total of 35 predictive models for dam failure modes were developed, designated as M1 through M35. The most accurate model, M11, achieved a 71% accuracy rate. This model excluded the volume variable, configured with three neighbors ( $k=3$ ), and a training/test sample size ratio of 90% to 10%. This model was subsequently applied to a database containing information on 10 Brazilian dams classified with emergency levels II and III by the National Mining Agency. Within this set, eight were identified as susceptible to credible failure due to slope instability, and two due to internal erosion.

**Keywords:** Geotechnical risk, Machine learning, K nearest neighbors, Tailings dams.



## LISTA DE FIGURAS

Figura 1 - Estrutura inicial de uma barragem pelo método de montante. ....	8
Figura 2 - Estrutura de uma barragem pelo método de montante. ....	9
Figura 3 - Estrutura inicial de uma barragem pelo método de jusante. ....	9
Figura 4 - Estrutura de uma barragem pelo método de jusante. ....	10
Figura 5 - Estrutura de uma barragem pelo método de linha de centro. ....	11
Figura 6 - Instabilidade de taludes por deslizamento. ....	12
Figura 7 - Ruptura causada por galgamento. a) Brecha inicial. b) Brecha intermediária. c) Ruptura total. ....	14
Figura 8 - I - Exemplo de erosão interna no maciço da barragem; ii - na fundação da barragem; iii - do maciço à fundação da barragem ....	15
Figura 9 - Sequência geral do processo da liquefação ....	16
Figura 10 - Processo de classificação. ....	17
Figura 11 - Modelo gráfico 2D do funcionamento do método KNN. ....	21
Figura 12 - Distância de Manhattan tanto em preto, quanto em verde entre os pontos A e B. Em azul a Distância Euclidiana entre os pontos A e B. ....	23
Figura 13 - Técnica de validação de subamostragem aleatória. ....	24
Figura 14 - Exemplo de matriz de confusão. ....	25
Figura 15 - Fluxograma do desenvolvimento do trabalho. ....	27
Figura 16 - Zonas de atividade sísmica do mundo. ....	29
Figura 17 - Mapa dos dez principais tipos de clima no mundo. ....	30
Figura 18 - Modelos preditivos gerados para cada configuração. ....	35
Figura 19 - Classificação do nível de emergência de uma barragem. ....	40
Figura 20 - Quantidade de barragens rompidas por cada modo de falha. ....	41

Figura 21 - Quantidade de barragens rompidas por zona de influência climática. .....	42
Figura 22 - Quantidade de barragens rompidas por zona de influência sísmica.	42
Figura 23 - Quantidade de barragens rompidas de acordo com o material de construção. .....	43
Figura 24 - Quantidade de barragens rompidas de acordo com o método de construção. .....	43
Figura 25 - Quantidade de barragens rompidas relacionadas a cada tipo de mineração. .....	44
Figura 26 - Quantidade de barragens por intervalos de altura. .....	45
Figura 27 - Quantidade de barragens por intervalos de volume (106 m <sup>3</sup> ). .....	45
Figura 28 - Erro associado a cada valor de K vizinhos. ....	47
Figura 29 - Variação da acurácia com alteração do estado aleatório. ....	47
Figura 30 - Variação da acurácia no intervalo de valores de K para C1, C3 e C7. .....	48
Figura 31 - Relação das variáveis volume e modo de falha. ....	49
Figura 32 - Relação das variáveis volume e clima. .....	50
Figura 33 - Relação das variáveis volume e risco sísmico. ....	50
Figura 34 - Relação das variáveis volume e altura. ....	51
Figura 35 - Relação das variáveis volume e método construtivo. ....	52

## LISTA DE TABELAS

Tabela 1 - Algoritmos de aprendizado de máquina.....	18
Tabela 2 - Codificação da variável modo de ruptura.....	30
Tabela 3 - Codificação da variável método construtivo.....	31
Tabela 4 - Codificação da variável risco sísmico. ....	31
Tabela 5 - Codificação da variável tipo de minério que gerou o rejeito armazenado. .....	31
Tabela 6 - Codificação da variável material de construção da barragem. ....	32
Tabela 7 - Codificação da variável clima. ....	33
Tabela 8 - Configurações adotadas para treinamento e geração de modelos preditivos.....	36
Tabela 9 - Barragens de rejeito com nível de emergência 2 e 3 no Brasil.....	39
Tabela 10 - Resultados dos modelos de previsão gerados. ....	46
Tabela 11 - Modo de falha crível para as barragens analisadas.....	52

## SUMÁRIO

1	Introdução .....	1
2	Objetivo .....	5
2.1	Objetivos específicos.....	5
3	Justificativa e relevância .....	6
4	Revisão Bibliográfica.....	7
4.1	Barragens de terra.....	7
4.2	Métodos Construtivos .....	8
4.2.1	Método de montante .....	8
4.2.2	Método de jusante.....	9
4.2.3	Método de linha de centro.....	10
4.3	Modos de Ruptura .....	11
4.3.1	Instabilidade de taludes .....	12
4.3.2	Instabilidade sísmica.....	13
4.3.3	Galgamento ou <i>overtopping</i> .....	13
4.3.4	Erosão interna ou <i>piping</i> .....	14
4.3.5	Liquefação.....	15
4.4	Aprendizado de Máquina.....	17
4.4.1	Aprendizado supervisionado .....	19
4.4.2	Aprendizado não supervisionado .....	19
4.4.3	K Nearest Neighbors – KNN .....	20
4.4.4	Métodos e métricas de validação .....	23
4.4.4.1	Subamostragem aleatória ( <i>Random subsampling</i> ) .....	24
4.4.4.2	Matriz de confusão.....	25

4.4.4.3	Métricas de validação .....	26
5	Metodologia.....	27
5.1	Banco de dados.....	28
5.2	Análise descritiva das variáveis .....	33
5.3	Treinamento de modelos preditivos.....	34
5.3.1	Pacotes <i>Python</i> utilizados .....	37
5.4	Avaliação e validação dos modelos preditivos gerados.....	38
5.5	Seleção do modelo preditivo .....	38
5.6	Aplicação do modelo preditivo selecionado .....	38
6	Resultados .....	41
6.1	Análise descritiva das variáveis .....	41
6.2	Acurácia dos modelos de predição gerados.....	45
6.3	Análise da influência das variáveis propostas .....	48
6.3.1	Justificativa para retirada da variável volume.....	49
6.4	Previsão do modo de falha crível das barragens em nível de emergência II e III no Brasil.....	52
7	Conclusão .....	54

## 1 INTRODUÇÃO

Segundo o Plano Nacional de Mineração (PNM, 2011), no Brasil o setor mineral é a base para diversas cadeias produtivas. Este setor participa com 4,2% do Produto Interno Bruto (PIB) e 20% do total das exportações brasileiras, gerando cerca de um milhão de empregos diretos, o equivalente a 8% dos empregos da indústria. Assim, devido a sua produção significativa de minério, o setor minerário também gera volumes consideráveis de rejeito em seu processo de beneficiamento (SOUZA JUNIOR; MOREIRA; HEINECK, 2018).

Para dispor o rejeito gerado após o beneficiamento do minério, a indústria mineral tradicionalmente realiza o lançamento da polpa de rejeito - mistura de água e sólidos - em barragem de contenção, construídas com material compactado em etapa única ou em mais etapas de alteamento (CARDOZO; PIMENTA; ZINGANO, 2016). Sendo assim, a construção e o desenvolvimento da estrutura devem se dar de forma contínua, permitindo o acompanhamento dos resultados, possíveis modificações e melhoramento do projeto inicial. Dessa forma, possibilita-se que todo o rejeito gerado na produção possa ser lançado e armazenado, de forma segura, minimizando os riscos de acidentes.

Porém, de acordo com Leão e Santiago (2022), diversos acidentes relacionados ao rompimento de barragens de rejeitos marcaram a trajetória da exploração mineral na história nacional e, mesmo com o conhecimento da legislação e de tecnologia diversas, as barragens continuam rompendo e causando prejuízos sociais, ambientais e econômicos nas áreas atingidas. Duarte (2008) explica que o fato de as barragens não serem executadas de acordo com critérios técnicos adequados pode contribuir para a ocorrência de acidentes desse tipo.

Nesse contexto, o ramo da ciência que estuda a estabilidade de tais estruturas é a geotecnia, que segundo Costa (2007) é “o ramo da Engenharia que agrupa as disciplinas científicas que se ocupam da caracterização e comportamento dos terrenos”, abrangendo as áreas de geologia de engenharia, mecânica dos solos e de rochas.

Segundo Gomes *et al.* (2005) a avaliação do risco deve começar pelo estudo de modos de falhas, para estimar as probabilidades e consequências associadas a cada um dos riscos identificados. Uma vez que, a depender do modo de falha, tem-se um diferente método para análise de estabilidade geotécnica da estrutura, o conhecimento prévio do modo de falha crível é importante para facilitar e fazer com que o gerenciamento de risco de barragem se torne mais eficiente e assertivo.

Aliado a esse gerenciamento, pode-se observar o avanço e uso cada vez mais disseminado da tecnologia, em diversas áreas, com o objetivo de melhorar identificação e análise de riscos geotécnicos. Diferentes softwares para análise de estabilidade de taludes e percolação de água se fazem presentes e ganham importância no gerenciamento de riscos em barragens.

Caputo (2014) afirma que é importante fazer a análise da estabilidade de talude de uma barragem, já que a maioria dos acidentes que acontecem se dá pela falha na estabilidade do maciço. Os mecanismos potenciais de ruptura devem ser previamente e adequadamente definidos, em função da geometria da barragem, das propriedades geotécnicas dos materiais do seu maciço e fundação, bem como da provável condição de fluxo de água ao longo dele.

No campo da tecnologia, importante na gestão de riscos como mencionado, a Inteligência Artificial e o Aprendizado de Máquina (*Machine Learning* - ML) tem ganhado cada vez mais visibilidade. Apesar de ser uma tecnologia recente, a sua utilização está amplamente difundida. A empresa de tecnologia Interop (2019), no mercado há 29 anos, cita como exemplo os aplicativos de transporte, como *Uber*, *Cabify*, entre outros, discutindo questões como: de que maneira eles determinam o preço dos passeios? Como minimizam o tempo de espera depois que um carro é solicitado? E como esses serviços combinam as rotas de maneira ideal com outros passageiros para minimizar desvios? A resposta para todas essas perguntas é a utilização de Aprendizado de Máquina.

A partir dos dados das corridas, a plataforma identifica padrões e se adapta conforme mudanças de comportamento ocorrem, ou seja, trabalham com a análise mais eficiente dos dados adquiridos, levando em consideração não apenas os dados

em si, mas também a correlação entre eles, algo que também é importante de se observar na análise de riscos geotécnicos. As máquinas não estão mais fazendo somente trabalhos manuais, mas também trabalhos racionais, tarefas que requerem o uso do que se considera inteligência (LUDERMIR, 2021).

Como exemplo do avanço de tal tecnologia no dia a dia, nota-se que as máquinas estão aprendendo a dirigir automóveis. Já existem carros que podem se mover sem motorista, como os carros do *Google* e da Tesla. Tais sistemas devem ser robustos o suficiente para lidar com situações para as quais não tenham sido treinados, assim como situações que possam colocar as vidas dos motoristas e pedestres em risco (FENG *et al.*, 2021).

Outra aplicação de sucesso que faz uso da inteligência artificial são os tradutores automáticos, como os tradutores do *Google*, que fazem traduções com precisão cada vez melhores. O processo de tradução de texto é feito com sucesso em razão da quantidade de textos, isto é, dados que existem disponíveis. A área da Inteligência Artificial que lida com a tradução e manipulação de textos é conhecida por Processamento de Linguagem Natural (TORFI *et al.*, 2020; OTTER *et al.*, 2021).

Na pesquisa apresentada neste trabalho, tem-se a apresentação da aplicação da técnica de Aprendizado de Máquina, *K Nearest Neighbor* (KNN) na análise de riscos geotécnicos na mineração, visando aumento de segurança e a melhoria dos resultados de análises que identificam probabilidades de falha.

A realização do estudo deu-se com base em uma abordagem teórico metodológica. Inicialmente, um banco de dados com 66 barragens de rejeito que romperam entre os anos 1928 e 2019 foi organizado e adaptado, com adição, adaptação e/ou padronização de variáveis. Em um segundo momento, técnica de análise multivariada de dados foi utilizada para elaboração de modelos preditores, posteriormente analisados quanto ao desempenho na predição dos modos de falha. Também se realizou a análise da influência que cada variável apresentava na acurácia dos modelos.

Com a definição do modelo de melhor desempenho, o mesmo foi utilizado para avaliar qual seria o modo de falha mais provável de ocorrer para um conjunto de 10



barragens de rejeito, classificadas com risco II ou III, em escala que vai de zero a três, sendo três o nível que indica ruptura iminente ou em andamento. Esse novo banco foi gerado com dados da Agência Nacional de Mineração – ANM, através do site de Classificação Nacional de Barragens de Mineração, considerando suas características estruturais e do ambiente onde se localizam.

Gerou-se, dessa forma, os resultados de predição, com base estatística e computacional, para as atuais barragens com maior risco de falha no Brasil, de acordo com a classificação da Agência Nacional de Mineração, alcançando o objetivo desse trabalho, em apresentar uma ferramenta capaz de fazer tal análise.

## **2 OBJETIVO**

Este trabalho objetiva estabelecer uma metodologia para proposição de um modelo preditivo, via aprendizado de máquina, para determinação do modo de falha crível para uma determinada barragem de rejeito de mineração.

### **2.1 Objetivos específicos**

- Determinar as variáveis que descrevem ou influenciam no fenômeno de ruptura em barragens de rejeito;
- Organizar um banco de dados adequado para aplicação das técnicas de aprendizado de máquina;
- Gerar modelos preditivos que possam ser aplicado na avaliação de risco de falha em barragens;
- Validar os modelos obtidos,
- Selecionar o modelo preditivo a ser proposto,
- Aplicar o modelo preditivo selecionado para previsão do modo de falha crível em barragens de rejeito que apresentam risco de se romperem.

### 3 JUSTIFICATIVA E RELEVÂNCIA

O desenvolvimento da pesquisa apresentada nesse trabalho buscou a criação de ferramenta para avaliação do modo de falha crível de barragens de rejeito utilizando técnicas de *aprendizado de máquina*. Com isso, objetivou-se reduzir a subjetividade do especialista das análises que, apesar do grande destaque dado na literatura aos estudos de análise e gestão de riscos de barragens de rejeito, ainda influencia de forma visível a tomada de decisões. Cabe-se ressaltar também as dificuldades associadas à quantificação dos parâmetros, que são facilmente identificadas.

Com a diminuição da subjetividade e quantificação das incertezas presentes nas análises de risco, pode-se chegar ao desenvolvimento de um sistema mais eficiente e confiável, que possa subsidiar decisões de engenharia e melhorar o conhecimento científico na gestão de riscos de barragens. Considerando, assim, a forte discussão sobre a necessidade de criação de métodos eficientes e facilitadores para tomada de decisões acerca de intervenções para manutenção e estabilização dessas estruturas ao longo do tempo, e que tais decisões estão sempre limitadas por condicionantes econômicos, levando à permanente necessidade de diminuição do grau de incerteza associado às análises de condição de estabilidade e suas consequências.

Além disso, cabe ressaltar que o desenvolvimento de ferramentas que subsidiem o gerenciamento de risco em barragens de rejeito é muito importante, uma vez que as duas últimas barragens que se romperam no estado de Minas Gerais, de acordo com a ANDES (2022) não estavam sequer classificadas como barragem de risco. Em vista disso, tal fato sugere a necessidade de proposição de métodos mais representativos e confiáveis para conclusão do nível de risco dado para uma determinada estrutura analisada.

## 4 REVISÃO BIBLIOGRÁFICA

### 4.1 Barragens de terra

De acordo com Cardozo, Pimenta e Zingano (2016) as barragens, juntamente com pilhas de estéril, são as maiores estruturas geotécnicas construídas pelo homem, sendo utilizadas pela humanidade há milênios para diversos fins, dentre eles: armazenamento de água ou resíduos da indústria, controle de vazões, geração de energia hidroelétrica, entre outros. Com o avanço da exploração mineral e a necessidade de conservação socioambiental relacionada com ela, o assunto barragem de rejeitos, que consistem em estruturas geotécnicas construídas para acondicionar os rejeitos provenientes da mineração, vem merecendo grande atenção, devido a seu potencial danoso em caso de falhas.

Historicamente, observa-se com certa frequência a ruptura de barragens de mineração e os danos causados por estas. Azam e Li (2010) apresentam um histórico mundial das falhas em barragens de mineração, em que cerca de 1,2% das barragens de mineração apresentaram algum tipo de falha, contra 0,01% das barragens civis, considerando os 100 anos anteriores à publicação.

Duarte (2008) salienta que barragens civis ou convencionais têm qualquer fim menos a contenção de rejeitos. Além disso, um ponto que também diferencia as barragens de rejeito, é a sua construção em etapas, as quais, acompanham a vida útil da mina. Desta forma, à medida que são gerados rejeitos, os alteamentos são executados. Essas estruturas também são reconhecidas por gerarem um impacto ambiental significativo, uma vez que eventuais falhas podem resultar em grandes volumes de rejeitos descarregados no meio ambiente (Duarte, 2008).

Diferente de barragens convencionais, que normalmente armazenam água, barragens de rejeito armazenam rejeitos dos processos minerais que podem variar de rejeitos granulares até solos de granulometria fina e alta plasticidade (lamas) (ARAUJO, 2006). Kossoff *et al.* (2014), definem os rejeitos destinados a barragens como mistura de rocha cominuída e fluidos do processo do beneficiamento mineral. Sendo assim, apresentam como característica física: granulometria fina e forma

angular, bem como composição química dependente da composição da rocha de origem e dos reagentes utilizados no processamento mineral.

## 4.2 Métodos Construtivos

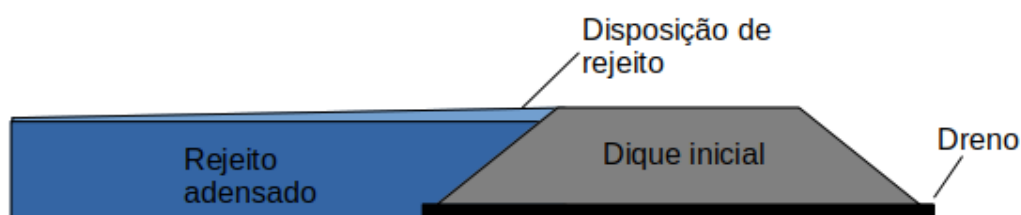
Sucintamente podem-se elencar três métodos aplicados na construção de estruturas de disposição de rejeito: método de montante, método de jusante e o método de linha de centro.

### 4.2.1 Método de montante

Tal método, também conhecido pela denominação *upstream*, pode ser considerado o mais simples e de mais baixo custo de construção. Esse método, considerado o mais perigoso (NOGUEIRA; SPRING; PLUMB, 2019), foi o utilizado nas estruturas que romperam mais recentemente no estado de Minas Gerais, nas cidades de Mariana e Brumadinho.

Segundo Martin e McRobert (1999) o método de alteamento a montante representa um desafio no âmbito geotécnico devido às tensões induzidas, potencial de liquefação e não consolidação do material utilizado como fundação. Sua construção inicia-se com um dique inicial utilizado como aterro compactado ou enrocamento, como mostra a Figura 1.

Figura 1 - Estrutura inicial de uma barragem pelo método de montante.

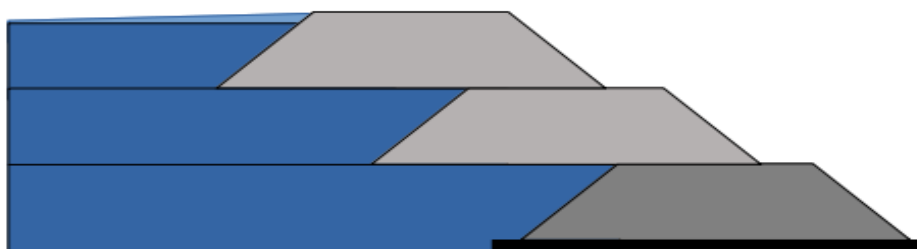


Fonte: O autor (2023).

O rejeito é descarregado hidráulicamente, por meio de canhões de hidrociclones, desde a parte mais alta, denominada crista, do dique de partida,

formando uma praia de rejeito. Por segregação natural a fração mais grossa fica próximo à crista, sequencialmente acontece o processo de adensamento deste material, que servirá como fundação para futuros diques de alteamento a serem construídos a partir da fração grossa dos rejeitos. O processo é repetido até que se alcance a cota máxima para alteamento prevista. Pode-se observar uma ilustração do método na Figura 2.

Figura 2 - Estrutura de uma barragem pelo método de montante.

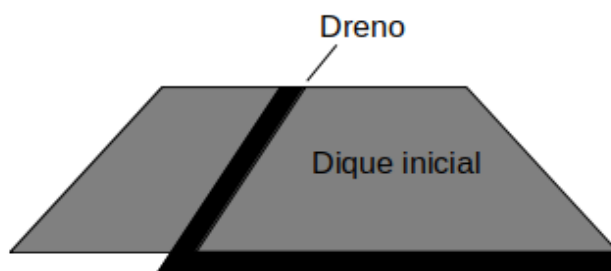


Fonte: O autor (2023).

#### 4.2.2 Método de jusante

O método de jusante, também chamado de *downstream*, de acordo com Carvalho *et al.* (2018), é o mais conservador entre os métodos citados, e foi desenvolvido para reduzir os riscos de liquefação em zonas de atividade sísmica. Da mesma maneira que o método de montante, é necessário a construção de um dique de partida, estrutura apresentada na Figura 3, e os rejeitos são depositados a montante dele.

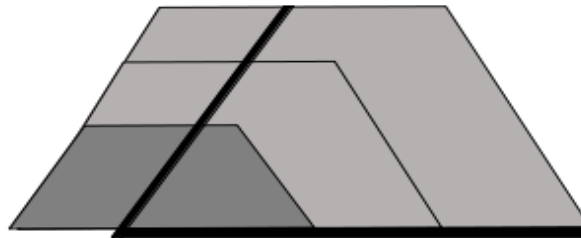
Figura 3 - Estrutura inicial de uma barragem pelo método de jusante.



Fonte: O autor (2023).

Conforme Campos (2022), ressalta-se que mesmo havendo um volume de água significativo, a estabilidade da estrutura pode não ser afetada devido sua fundação ser em terreno natural, além de haver a instalação de zonas de drenagem e núcleos impermeáveis. A desvantagem desta alternativa é o alto custo dos alteamentos, provenientes do grande volume de aterro necessário. O limite de altura de uma barragem de rejeitos alteada para jusante dependerá, basicamente, da área de terreno disponível. A Figura 4, apresenta uma representação do método.

Figura 4 - Estrutura de uma barragem pelo método de jusante.



Fonte: O autor (2023).

De acordo com Chammas (1989, apud PASSOS, 2009) as vantagens do método de jusante estão na sua resistência a carregamentos dinâmicos. Isto deve-se ao fato de escalonar a construção sem interferir na segurança, dessa forma facilita a drenagem, possui baixa susceptibilidade de liquefação e simplicidade na operação.

#### 4.2.3 Método de linha de centro

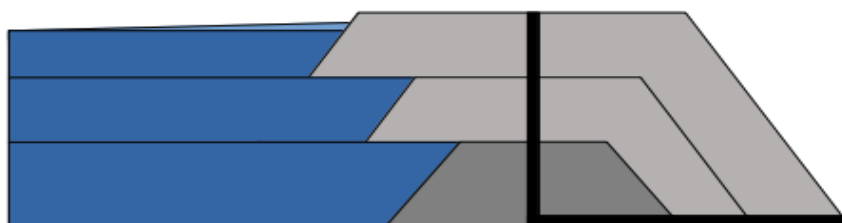
Tal metodologia de construção, denominada em inglês como *centerline*, possui maior estabilidade do que uma barragem alteada para montante e não requer um grande volume de materiais como no alteamento para jusante. O método da linha de centro é um método intermediário que tenta minimizar as desvantagens entre o método de montante e o de jusante (PASSOS, 2009).

Da mesma forma que os métodos anteriores, precisa ser construído um dique de partida com o intuito de formar a praia de rejeitos a montante. A disposição dos rejeitos

é feita com o lançamento a partir da crista do dique inicial, assim como no método de alteamento a montante.

Quando é preciso fazer novos alteamentos, são construídos novos diques, tanto sobre os rejeitos a montante quanto sobre o aterro do dique anterior, mantendo o eixo de simetria inalterado. Na Figura 5 é apresentado um esquema desse método construtivo.

Figura 5 - Estrutura de uma barragem pelo método de linha de centro.



Fonte: O autor (2023).

### 4.3 Modos de Ruptura

Diversas são as implicações que o ambiente, a ação antrópica e a falta de monitoramento adequado podem causar no que diz respeito à ruptura de barragens. Aspectos como alturas e volumes cada vez mais elevados, contribuem para a necessidade de uma atenção redobrada, uma vez que aumenta a dificuldade na construção dos barramentos à medida em que mais variáveis devem ser levadas em consideração, como, por exemplo, fatores geotécnicos, atmosféricos e ambientais.

Ademais, para uma boa gestão e garantia de segurança das estruturas, o estudo e implementação de um eficiente sistema de drenagem é indispensável. A instalação de drenos internos é comum para todos os métodos construtivos de barragens e sua construção é realizada com o objetivo de prevenir excessos de poropressão, favoráveis ao movimento, devido ao fluxo interno de água na barragem.

Dentre os diversos casos de rupturas de barragens na história da mineração, alguns modos de falha foram estudados e mais bem compreendidos pela comunidade geotécnica. Os mais comuns são: ruptura por instabilidade de taludes; liquefação;



galgamento ou *overtopping*; erosão interna ou *piping*; bem como instabilidade devido a terremotos, erosão externa ou até mesmo inadequações estruturais de projeto.

#### 4.3.1 Instabilidade de taludes

Os deslizamentos ou escorregamentos são problemas comuns em barragens de terra, pois o material que compõe o talude naturalmente tende a escorregar devido principalmente à força da gravidade, que é suportada pela resistência ao cisalhamento do material (DYMINSKI, 2014). Segundo Caputo (1987), as principais causas de rupturas em barragens de terra são o galgamento, as infiltrações e os deslizamentos, o qual se caracteriza pelo movimento descendente de uma massa de solo em uma superfície de ruptura, como mostra a Figura 6.

As variações nas condições de poropressão, carregamento e variações nas forças de cisalhamento do aterro resultam em mudanças nos fatores de segurança (razão entre forças de resistência ao cisalhamento e forças solicitantes) da estrutura.

Figura 6 - Instabilidade de taludes por deslizamento



Fonte: Das e Sobhan (2014).

Como consequência, muitas vezes é necessário realizar análises de estabilidade correspondentes a várias condições que refletem diferentes etapas da construção de uma barragem, uma vez que o fator de segurança dessas estruturas

pode aumentar ou diminuir ao longo de sua vida útil e da condição de contorno e carregamento da estrutura.

#### **4.3.2 Instabilidade sísmica**

Tremores submetem uma barragem a variações cíclicas de carregamento em um curto intervalo de tempo, podendo causar instabilidades ou deformações permanentes nos taludes, dependendo da gravidade do tremor e do seu efeito sobre a resistência do aterro (DUNCAN *et al.*, 2014). Ladeira (2007) afirma que esses abalos sísmicos também podem liquefazer os solos do maciço da barragem, sendo função da frequência de ocorrência e da magnitude do tremor. Segundo Seed (1979), deve-se inclusive considerar a possibilidade de acontecerem falhas retardadas (aquelas que ocorrem mesmo após os movimentos do terremoto terem cessado) devido à redistribuição da pressão da água nos poros dentro de um aterro.

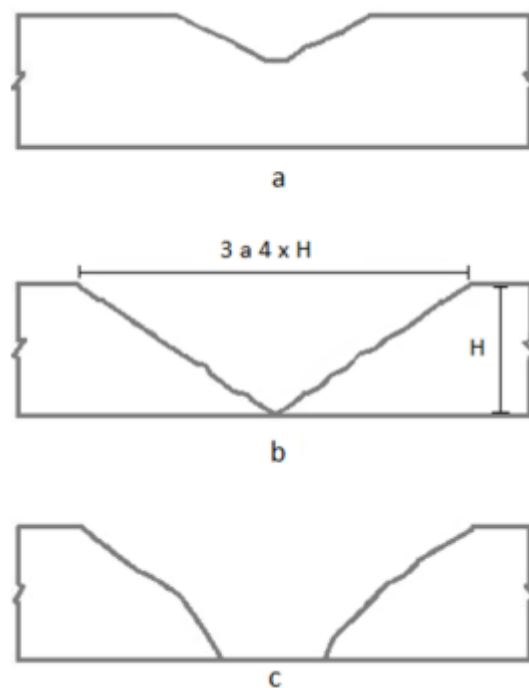
Possíveis maneiras que um terremoto pode causar a falha de uma barragens de aterro compactado são: rompimento da barragem por movimento de falha principal na fundação; perda de bordo livre devido a movimentos tectônicos diferenciais; falhas de declive induzidas por movimentos do solo; perda do bordo livre devido a falhas de declividade ou compactação do solo; deslizamento da barragem em materiais fracos da fundação; *piping* através de rachaduras induzidas por movimentos do solo; galgamento da barragem devido deformação da crista e falha do vertedouro ou outro dispositivo de descarga (SEED, 1979).

#### **4.3.3 Galgamento ou *overtopping***

Segundo Balbi (2005), o modo de falha por galgamento acontece quando o nível de água do reservatório aumenta a ponto de ultrapassar a cota máxima da barragem. Esse fenômeno normalmente acontece quando chuvas intensas promovem uma alta elevação no nível de água, gerando vazões maiores do que a capacidade do vertedouro e provocando, assim, a destruição da barragem.

Assim como as propriedades do material e as características do aterro, a duração do galgamento é um dos fatores mais determinantes na dimensão das brechas de ruptura (JONATAS, 2013). Segundo Johnson e Illes (1976), em barragens de aterro, o início da brecha advém com uma forma triangular, em que, com a saída da água pela superfície da barragem, são erodidos os materiais menos resistentes. Caso sejam mantidos os níveis elevados da cheia, a geometria tende a tomar uma forma trapezoidal, sendo a extensão dessa erosão nas laterais da brecha dependente da dimensão do reservatório e do volume de água armazenado. A Figura 7 ilustra tal situação.

Figura 7 - Ruptura causada por galgamento. a) Brecha inicial. b) Brecha intermediária. c) Ruptura total.



Fonte: Johnson e Illes (1976).

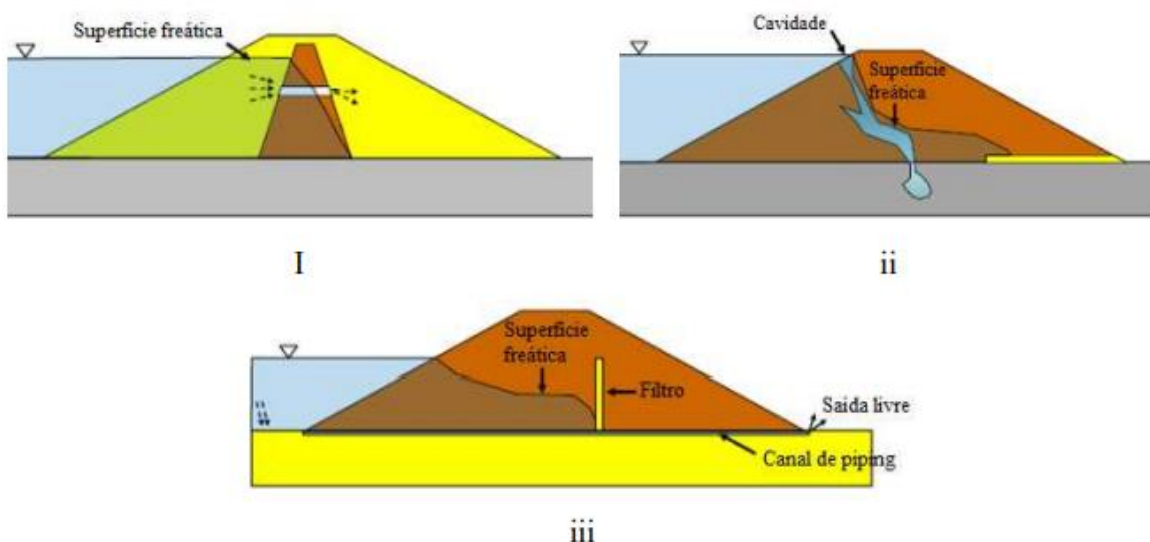
#### 4.3.4 Erosão interna ou *piping*

As rupturas por erosão interna ou *piping* são originadas pelo carreamento de partículas de solo devido à percolação da água através da barragem (FEMA, 2015).

Este modo de falha pode ocorrer quando há uma falha no sistema de drenagem interno em direcionar o fluxo na barragem, que provoca a remoção de partículas do interior do aterro, formando “tubos” vazios, causando colapsos e escorregamentos laterais do terreno, como pode-se observar, em etapas sucessivas, na Figura 8.

De acordo com Marques Filho e Geraldo (2002), este fenômeno ocorre em solos pouco coesos onde o gradiente hidráulico é maior que a coesão do solo. Rafael e Romanel (2014) afirmam que a velocidade de disposição de rejeitos e de alteamentos junto com drenagem interna ineficiente contribuem para ocorrência de *piping* em barragens com método construtivo a montante.

Figura 8 - I - Exemplo de erosão interna no maciço da barragem; ii - na fundação da barragem; iii - do maciço à fundação da barragem



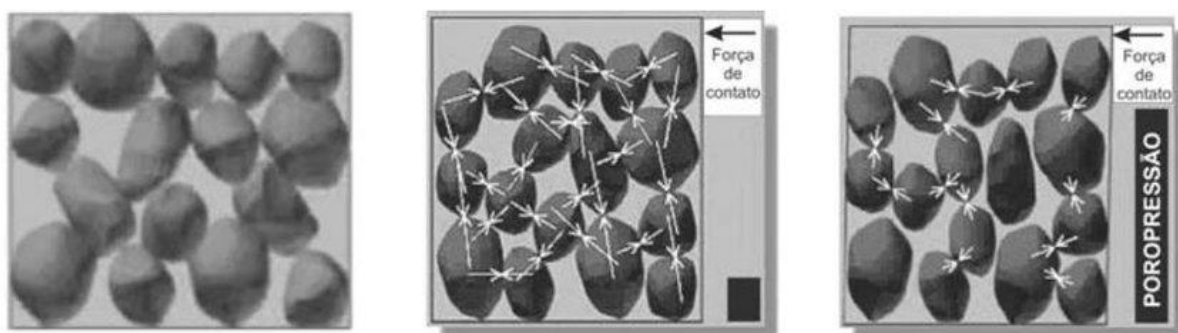
Fonte: USACE (2015).

#### 4.3.5 Liquefação

A liquefação é um processo de ruptura que ocorre em solos não coesivos, saturados e em condições de carregamento não drenado. Durante a aplicação do carregamento, a poropressão aumenta até um valor igual à tensão de confinamento inicial, a tensão efetiva ou intergranular existente no esqueleto do material é reduzida

a zero e, em consequência, o material perde praticamente toda sua resistência cisalhante, comportando-se como um líquido viscoso (CASAGRANDE, 1975; CASTRO, 1969; POULOS, 1981; TERZAGHI; PECK; MESRI, 1996). Durante a liquefação a posição relativa dos grãos está em constante mudança, de forma que é mantida uma resistência mínima, como pode-se observar na Figura 9.

Figura 9 - Sequência geral do processo da liquefação



Fonte: Pereira (2005).

Como característica importante desse modo de falha pode-se citar o fato de que esse tipo de ruptura ocorre em certas regiões da massa de solo e não apenas ao longo de uma determinada superfície de falha, provocando grandes deformações e múltiplos mecanismos de ruptura (FREIRE NETO, 2009; GUILLÉN, 2008; LOAYZA, 2009).

Segundo Maturano (2012), a liquefação de solos granulares é observada em depósitos de solos naturais, aterros e depósitos de rejeitos de mineração sob carregamento estático e sob carregamento dinâmico. O autor apresenta também possibilidade de sua ocorrência como efeito de vibrações causadas pela cravação de estacas, por tráfego de veículos, exploração geofísica, explosões, além de vibrações sísmicas, levando também à instabilização de taludes.

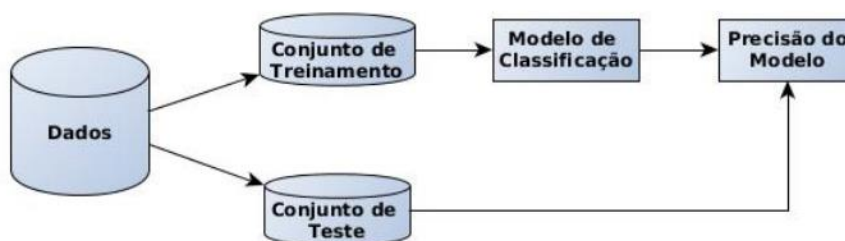
Em vista disso, tal modo de falha, neste trabalho, foi incorporado ao modo de falha instabilidade de taludes e ao de ruptura por abalo sísmico, a depender das características da barragem analisada.

#### 4.4 Aprendizado de Máquina

De acordo com Fontana (2020) o aprendizado de máquina é um termo geral utilizado para definir uma série de algoritmos que extraem informação a partir de um conjunto de dados, sem ser necessário definir um modelo matemático específico. Han, Kamber e Pei (2011) descrevem a abordagem geral para a classificação como um processo de duas etapas, como pode-se observar na Figura 10.

Inicialmente, é construído um modelo de classificação, através do conjunto de treinamento. Após, através do conjunto de teste, determina-se se a precisão do modelo é aceitável, e se assim for, o modelo é usado para classificar novos dados. O objetivo principal é desenvolver modelos e sistemas que possam identificar padrões nos dados e fazer previsões ou tomar decisões com base nesses padrões.

Figura 10 - Processo de classificação.



Fonte: Adaptado de Han, Kamber e Pei (2011)

Existem várias abordagens e algoritmos utilizados no aprendizado de máquina. Por exemplo, algoritmos de regressão linear são frequentemente usados para modelar relacionamentos lineares entre variáveis, enquanto árvores de decisão podem ser usadas para tomar decisões com base em condições lógicas. Redes neurais artificiais, inspiradas no funcionamento do cérebro, são algoritmos complexos que podem aprender a partir de dados para realizar tarefas como classificação e reconhecimento de padrões.

A tabela 1 apresenta, de acordo com Brunton, Noack e Koumoutsakos (2020), exemplos de tais abordagens e algoritmos, de acordo com diferentes grupos.

Tabela 1 - Algoritmos de aprendizado de máquina.

Classificação	Máquinas de vetor suporte; Árvores de decisão; Florestas aleatórias (random forests); Redes neurais artificiais; K-nearest neighbor.
Regressão	Linear; Linear generalizado; Processo Gaussiano.
Otimização e Controle	Controle linear; Algoritmos genéticos; Controle preditivo de modelo profundo; Estimação de algoritmos de distribuição; Estratégias evolutivas.
Aprendizado por reforço	Q-learning; Processo de decisão de Markov; Aprendizado por reforço profundo.
Modelos generativos	Redes adversárias generativas.
Agrupamento	K-means; Agrupamento espectral.
Redução de dimensionalidade.	POD/PCA; Codificador automático; Mapas autoorganizáveis; Mapas de difusão.

Fonte: Adaptado de Brunton, Noack e Koumoutsakos (2020).

Dentre as referências científicas que fornecem uma base sólida para o aprendizado de máquina pode-se citar Hastie, Tibshirani & Friedman (2009), que fazem uma introdução abrangente aos fundamentos do aprendizado de máquina, abordando tópicos como regressão, classificação, redução de dimensionalidade e seleção de modelos. Além desses autores, Bishop (2006) explora os princípios e técnicas do aprendizado de máquina, abrangendo conceitos estatísticos e probabilísticos, redes neurais, máquinas de vetores de suporte e outros algoritmos.

Essas referências, bem como outros trabalhos, permitem uma compreensão mais aprofundada dos algoritmos e das técnicas utilizadas nesse campo. No entanto, vale ressaltar que o campo do aprendizado de máquina está em constante evolução,

e novas pesquisas e avanços são publicados regularmente em conferências e revistas científicas especializadas, como a "*Journal of Machine Learning Research*" e a "*Neural Information Processing Systems (NeurIPS)*".

#### **4.4.1 Aprendizado supervisionado**

De acordo com Mitchell (1997) o aprendizado supervisionado consiste em uma abordagem em que um modelo é treinado usando um conjunto de dados rotulados, ou seja, dados nos quais as respostas corretas são conhecidas. O objetivo é aprender um mapeamento entre as entradas dos dados e as saídas desejadas por meio de algoritmos de aprendizado.

Durante o treinamento supervisionado, o modelo é exposto a exemplos de treinamento que consistem em pares de entrada-saída. Com base nesses exemplos, o modelo ajusta seus parâmetros internos para aprender a relação entre essas entradas e saídas. Posteriormente, o modelo pode ser usado para fazer previsões ou tomar decisões em novos dados de entrada.

#### **4.4.2 Aprendizado não supervisionado**

Muito comumente não se tem os rótulos dos dados em análise, ou seja, não se tem informações *a priori* sobre as classes as quais os dados pertencem, possuindo, em problemas desse tipo, apenas um conjunto de vetores de características, a partir do qual busca-se revelar as similaridades e agrupar vetores parecidos. Tal abordagem é conhecida como aprendizado não supervisionado (THEODORIDIS; KOUTROUMBAS, 2009) (DUDA; HART; STORK; 2001).

Diferentemente do aprendizado supervisionado, em que há uma resposta conhecida para cada exemplo, no aprendizado não supervisionado o modelo explora a estrutura intrínseca dos dados para identificar agrupamentos, similaridades ou outras características relevantes.



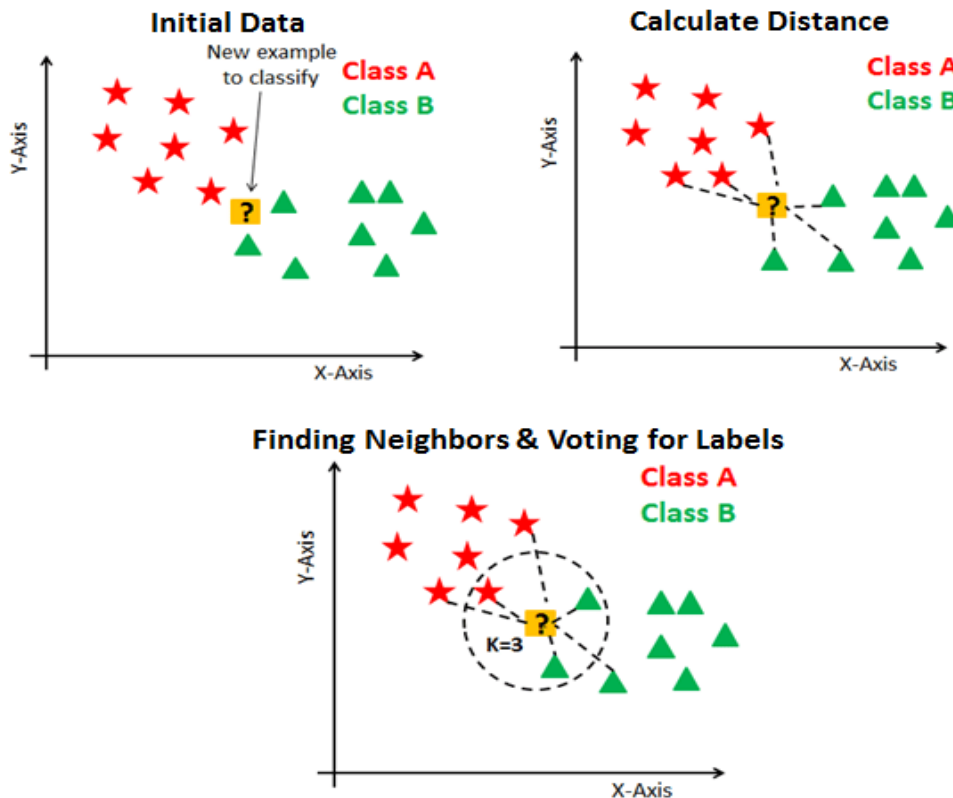
Existem várias técnicas e algoritmos de aprendizado não supervisionado, como clusterização (por exemplo, k-means e agrupamento hierárquico), redução de dimensionalidade (como análise de componentes principais - PCA) e associação de regras (por exemplo, algoritmos Apriori). Wu *et al.* (2007) apresentam uma visão geral dos dez algoritmos mais influentes em análises de dados, onde incluem técnicas não supervisionadas, como a clusterização e análise de componentes principais.

#### **4.4.3 K Nearest Neighbors – KNN**

Conforme Han *et al.* (2011) o KNN “é um dos algoritmos de aprendizado de máquina mais simples e amplamente utilizados para classificação e regressão. Ele tem a vantagem de ser não paramétrico, o que significa que não faz suposições específicas sobre a distribuição dos dados. Além disso, o KNN é fácil de implementar e interpretar, tornando-o uma escolha popular em várias aplicações.” Ele se baseia no princípio de que objetos similares tendem a estar próximos uns dos outros no espaço de características. A classificação ou previsão é feita com base na maioria dos votos dos K vizinhos mais próximos.

A Figura 11 mostra um exemplo de funcionamento do algoritmo KNN. Nota-se que existem 2 classes diferentes classificadas nos eixos x e y, e existe uma incógnita no centro. Inicialmente é calculada a distância entre essa incógnita e os demais pontos, a partir disso, é predefinido pelo usuário o valor de K, que vai delimitar a quantidade de dados analisados. No exemplo dado, o valor de K escolhido é 3, essa é a quantidade de vizinhos mais próximos a serem considerados para a classificação da incógnita. A classificação ocorrerá com base no voto majoritário. Neste caso a incógnita, considerando k igual a 3, será classificada no grupo dos triângulos verdes.

Figura 11 - Modelo gráfico 2D do funcionamento do método KNN.



Fonte: Azank (2019).

Uma das principais vantagens do KNN é a simplicidade de implementação e interpretação. Além disso, ele pode ser aplicado a dados não lineares e é robusto a outliers. No entanto, o desempenho do KNN pode ser afetado pelo número de vizinhos escolhidos (valor de K) e pela dimensionalidade dos dados (RAPOSO, 2021).

Conforme Hair *et al.* (2018) as distâncias mais comumente usadas no contexto do algoritmo K Nearest Neighbors (KNN) e em outros métodos de aprendizado de máquina incluem a distância euclidiana e a distância de Manhattan. Ambas as métricas são utilizadas para calcular a distância entre dois pontos em um espaço de características.

A distância euclidiana, também conhecida como norma euclidiana ou distância L2, é calculada como a medida da linha reta entre dois pontos em um espaço euclidiano. Segundo Prado (2015), essa métrica considera a raiz quadrada do

somatório dos quadrados das diferenças entre os valores de atributos para cada par de variáveis, conforme a equação (1).

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} \quad (1)$$

Nessa equação,  $d(x, y)$  representa a distância entre os pontos  $x$  e  $y$ ,  $(x_1, x_2, \dots, x_n)$  e  $(y_1, y_2, \dots, y_n)$  representam as coordenadas dos pontos  $x$  e  $y$  em um espaço de  $n$  dimensões.

De acordo com Cordeiro (2012), a distância euclidiana é considerada a métrica ideal pelo fato de adotar as medidas reais dos espaços euclidianos, apresentando ótimos resultados na maioria das aplicações.

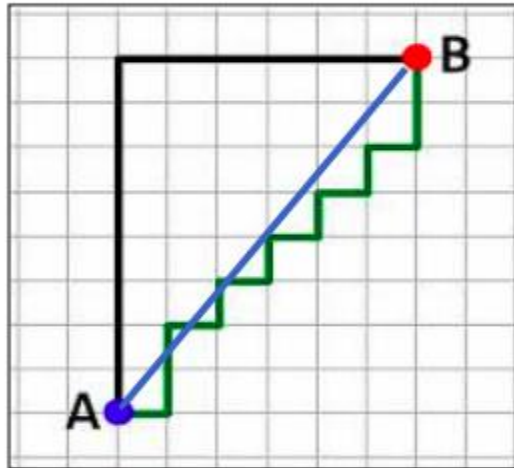
De acordo com Lei *et al.* (2017), a distância de Manhattan, também chamada *City Block*, ou Geometria do Táxi, é uma medida de cálculo de distância que ao contrário da distância Euclidiana não se eleva as diferenças ao quadrado e nem se obtém a raiz, permitindo que uma considerável distância entre dois pontos únicos não tenha uma influência elevada no valor de distância final. A fórmula para a distância de Manhattan é dada pela equação (2).

$$d(x, y) = |x_1 - y_1| + |x_2 - y_2| + \dots + |x_n - y_n| \quad (2)$$

A distância de Manhattan é menos sensível a outliers e é mais adequada quando as características têm escalas diferentes ou quando as relações de dependência entre elas são menos importantes.

Souza *et al.* (2012) dizem que essa métrica possui este nome em referência às ruas da ilha de Manhattan, em virtude da disposição quadriculada. Sendo assim, a distância de Manhattan foi utilizada para resolver uma situação que envolve qual o percurso a ser realizado por um carro entre dois pontos da cidade. Considerando que o carro deve andar somente em ruas de pavimento asfáltico. Tal situação é exemplificada na Figura 12.

Figura 12 - Distância de Manhattan tanto em preto, quanto em verde entre os pontos A e B. Em azul a Distância Euclidiana entre os pontos A e B.



Fonte: Souza *et al.* (2012).

Em suma, ambas as distâncias, euclidiana e de Manhattan, podem ser utilizadas no algoritmo KNN para calcular as distâncias entre os pontos de treinamento e o novo exemplo a ser classificado. A escolha entre essas métricas depende do domínio do problema, da natureza dos dados e do impacto relativo das diferentes dimensões nas características.

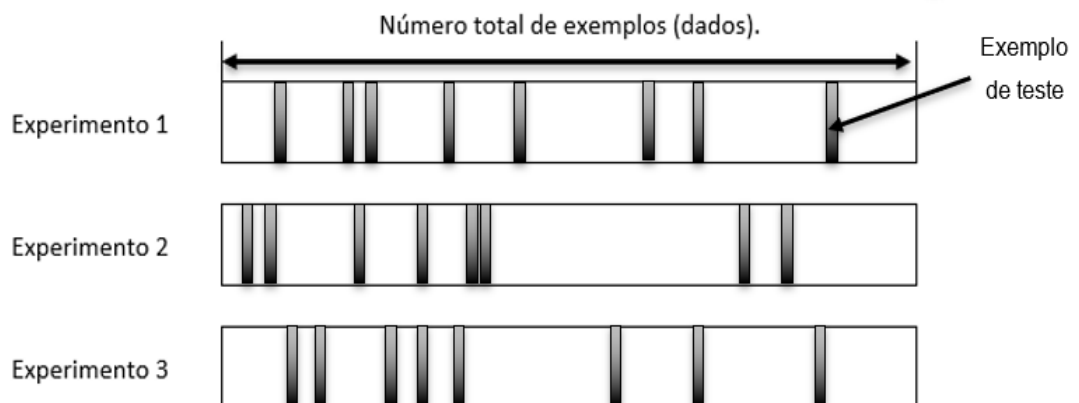
#### 4.4.4 Métodos e métricas de validação

É fundamental em problemas onde o objetivo da modelagem é a predição, a utilização de métodos de validação, que em suma buscam estimar o quão acurado e generalizável é um modelo na prática, ou seja, o seu desempenho em um novo conjunto de dados. Dentre os métodos utilizados para validação de modelos pode-se citar o método de subamostragem aleatória, também conhecida como validação cruzada de Monte Carlo. Entre as métricas de validação pode-se citar a probabilidade global de acerto (acurácia) e a taxa de erro aparente.

#### 4.4.4.1 Subamostragem aleatória (*Random subsampling*)

De acordo com Kumar (2018), nesta técnica vários conjuntos de dados são escolhidos aleatoriamente do conjunto de dados e combinados para formar um conjunto de dados de teste. Os dados restantes formam o conjunto de dados de treinamento. O diagrama a seguir representa a técnica de validação de subamostragem aleatória. Uma representação do método é apresentada na Figura 13.

Figura 13 - Técnica de validação de subamostragem aleatória.



Fonte: Adaptado de Kumar (2018).

A taxa de erro do modelo é a média da taxa de erro de cada iteração. A equação (3) mostra o cálculo do erro global como a média do erro de cada iteração.

$$Erro = \frac{1}{3} \sum_{i=1}^3 Erroi \quad (3)$$

De acordo com Dieterle (2019) o particionamento aleatório dos dados pode ser repetido com frequência arbitrária. Em contraste com um procedimento completo de validação cruzada, a subamostragem aleatória tem se mostrado assintoticamente consistente, resultando em previsões mais pessimistas dos dados de teste em

comparação com a validação cruzada. As previsões dos dados de teste fornecem uma estimativa realista das previsões dos dados de validação externa.

#### 4.4.4.2 Matriz de confusão

Antes de abordar as métricas de validação, é necessário compreender o conceito de matriz de confusão. De acordo com Godbole & Sarawagi (2004), uma matriz de confusão para um problema de  $n$  classes é uma matriz  $M$  de ordem  $n \times n$ , no qual o elemento  $M_{ij}$  corresponde ao número de classificações corretas para a classe  $i$ , quando  $i = j$ . Para  $i \neq j$ , o elemento  $M_{ij}$  corresponde ao número de classificações da classe  $i$  que foram definidas incorretamente como classe  $j$ , e vice-versa.

Conforme Mariano (2021), o número de acertos para cada classe se localiza na diagonal principal da matriz  $M$ . Usualmente podem ser depreendidas algumas métricas de avaliação do modelo de classificação apenas com base na sua matriz de confusão. A Figura 14 apresenta uma matriz de confusão para o caso de variável categórica com 2 populações, sendo o eixo horizontal a condição predita e o vertical a condição verdadeira.

Figura 14 - Exemplo de matriz de confusão.

	Positivo	Negativo
Positivo (T)	Verdadeiro Positivo (TP)	Falso Negativo (FN)
Negativo (N)	Falso Positivo (FP)	Verdadeiro Negativo (TN)
	Valor Predito Positivo (PPV)	Valor Predito Negativo (NPV)

Fonte: Maiorano (2021).

#### 4.4.4.3 Métricas de validação

Sejam  $n_1$  e  $n_2$  o número de elementos amostrais provenientes das populações  $\pi_1$  e  $\pi_2$ , respectivamente. Sendo  $n_{11}$  e  $n_{22}$  o número de elementos amostrais provenientes das populações  $\pi_1$  e  $\pi_2$  que foram classificados corretamente pela regra de discriminação, ou seja, todos os elementos representados, respectivamente, por TP e TN da Figura 13. A equação (4) define a probabilidade global de acerto (acurácia).

$$PGA = \frac{(n_{11} + n_{22})}{(n_1 + n_2)} \quad (4)$$

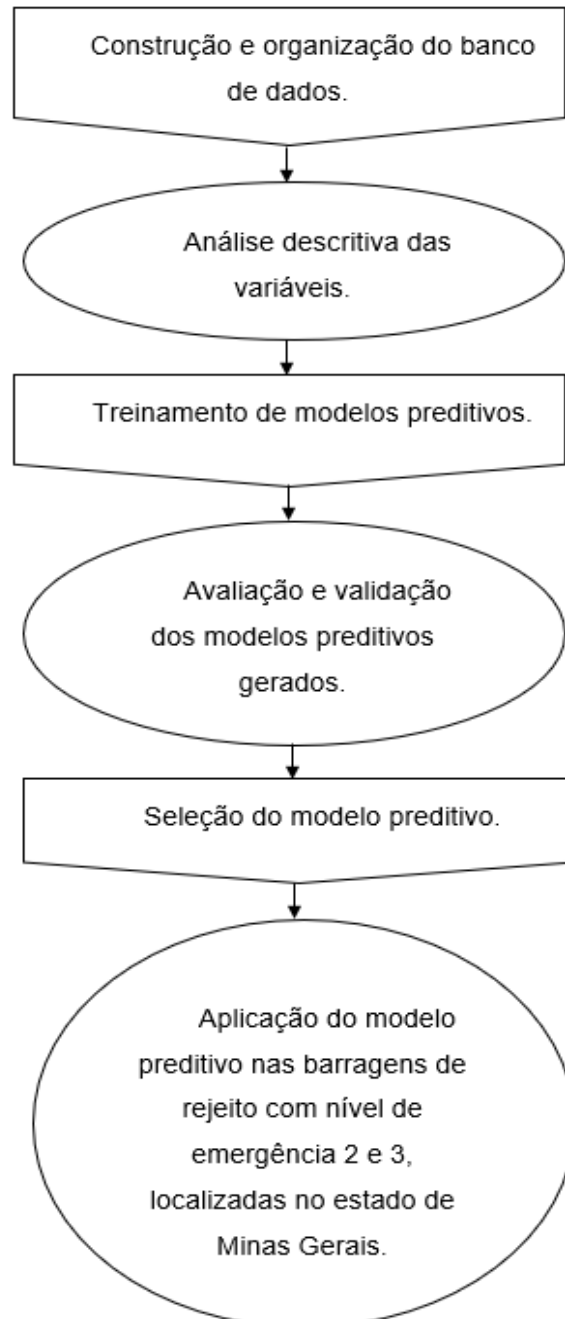
A taxa de erro aparente é a proporção de elementos classificados equivocadamente, ou seja,  $n_{12}$  (elementos da população 2 classificados como da população 1) e  $n_{21}$  (elementos da população 1 classificados como da população 2), representados, respectivamente, por FN e FP na Figura 13. Sendo assim a taxa de erro aparente pode ser definida, na equação (5).

$$TEA = \frac{(n_{12} + n_{21})}{(n_1 + n_2)} \quad (5)$$

## 5 METODOLOGIA

Na Figura 15 têm-se um fluxograma indicando a metodologia geral aplicada na pesquisa.

Figura 15 - Fluxograma do desenvolvimento do trabalho.



Fonte: O autor (2023).



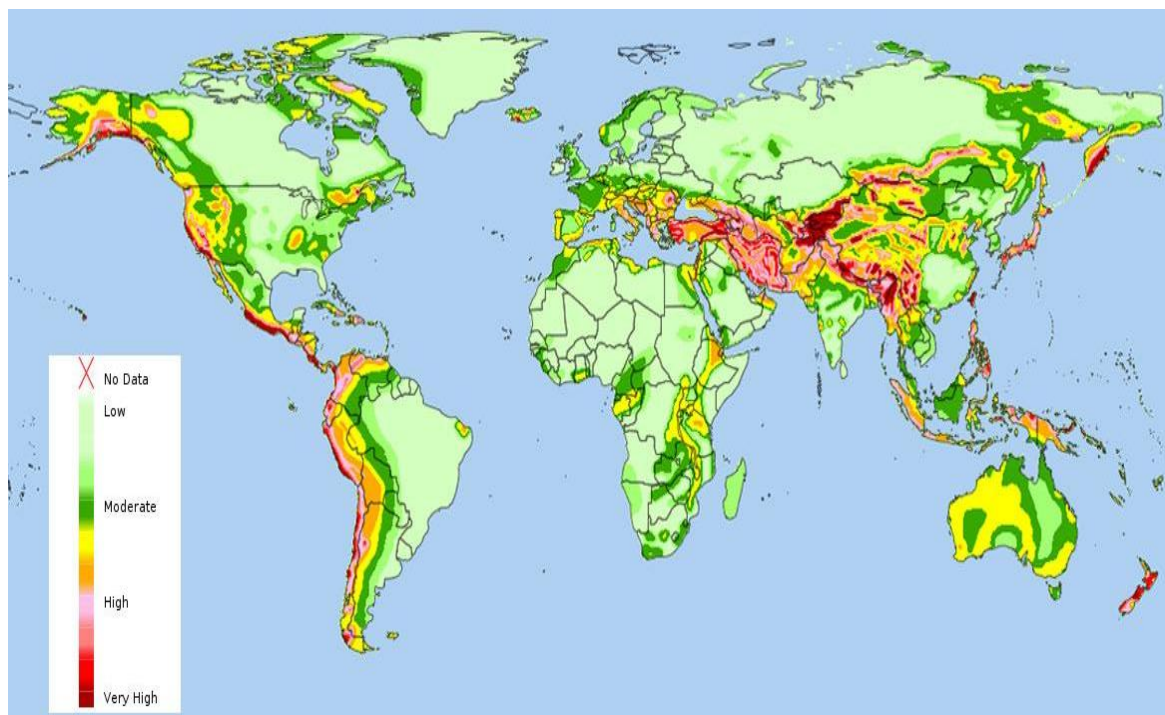
## 5.1 Banco de dados

O banco de dados utilizado neste trabalho foi construído, tendo como base um outro banco com 350 barragens rompidas de todo o mundo, obtido por meio do site <https://worldminetailingsfailures.org/>, bem como informações obtidas em outros meios de pesquisa. Porém, como muitas das informações estavam incompletas, foram retiradas as barragens que não apresentavam dados suficientes para análise, resultando em um banco final composto por 66 barragens, que é apresentado no Apêndice A, de forma codificada para aplicação na análise via KNN.

Para cada elemento amostral (barragem) do banco de dados, é conhecido: minério que gerou o rejeito armazenado, material de construção da barragem, método construtivo, altura, volume e o modo de falha. Além dessas variáveis, foi realizado o acréscimo de outras duas: o risco sísmico do local onde a barragem se encontrava e o clima predominante na região.

A variável risco sísmico foi definida considerando uma escala de classificação que vai de risco muito baixo até muito alto, por meio de um mapa (Figura 16), com informações do Centro de Dados e Aplicações Socioeconômicas – SEDAC. O SEDAC é um sistema de dados e informações do sistema de observação da Terra da NASA (EOSDIS), hospedado pelo Centro para Rede Internacional de Informações sobre Ciências da Terra da Universidade de Columbia – EUA.

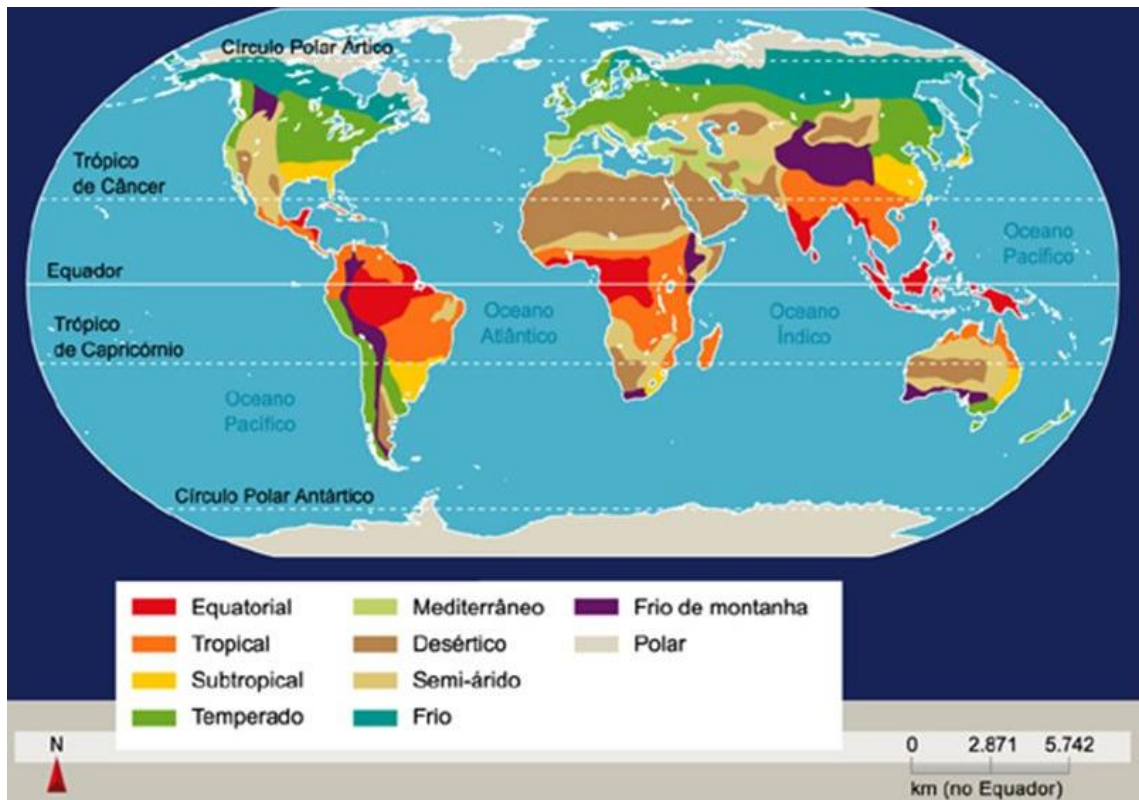
Figura 16 - Zonas de atividade sísmica do mundo.



Fonte: SEDAC (2014).

Quanto à variável clima, foi utilizado o mapa apresentado na Figura 17. Nele são estratificados os climas de acordo com classificação climática de Koppen, que segundo Rolim *et al.* (2007) é a mais utilizada no mundo. Ela se baseia em zonas de vegetação.

Figura 17 - Mapa dos dez principais tipos de clima no mundo.



Fonte: Freire (2023).

Por se tratar de variáveis qualitativas, houve a necessidade de codificação do minério que gerou o rejeito armazenado, material de construção da barragem, método construtivo e o modo de falha, bem como do risco sísmico e clima. Já para altura e volume, que são quantitativas, seus valores absolutos foram mantidos no banco.

As tabelas de 2 a 7 indicam como foram codificadas as informações referentes a cada variável.

Tabela 2 - Codificação da variável modo de ruptura.

Modo de Ruptura	Código
Instabilidade de talude	1
Erosão interna ( <i>piping</i> )	2
Fundação	3

Galgamento	4
Inadequações estruturais	5
Abalo sísmico	6
Erosão externa	7

Fonte: O autor (2023).

Tabela 3 - Codificação da variável método construtivo.

Método construtivo	Código
Montante	1
Linha de centro	2
Jusante	3
Etapa única	4
Montante e Jusante	5

Fonte: O autor (2023).

Tabela 4 - Codificação da variável risco sísmico.

Risco sísmico	Código
Muito baixo	1
Baixo	2
Médio	3
Alto	4
Muito alto	5

Fonte: O autor (2023).

Tabela 5 - Codificação da variável tipo de minério que gerou o rejeito armazenado.

Minério que gerou o rejeito armazenado	Código
Fe	1
Al	2
Cu	3
Au	4

Cu Au	5
Pb Zn	6
Ag Pb	7
Mo	8
Coal	9
Sn	10
U	11
Pt	12
Zn Pb Cu	13
Au Ag	14
Areia	15
Limestone	16
Mica	17
Ag	18
Gypsum	19
Pb Zn Cu Ag	20

Fonte: O autor (2023).

Tabela 6 - Codificação da variável material de construção da barragem.

Material de construção da barragem	Código
Rejeito	1
Ciclonado	2
Estéril	3
Aterro	4
Enrocamento	5
Cimento ( <i>fly-ash</i> )	6
Retenção de água	7
Estéril e Aterro	8
Aterro e Rejeito	9
Argilito, aleurolito	10

Fonte: O autor (2023).

Tabela 7 - Codificação da variável clima.

Clima	Código
Semiárido	1
Equatorial	2
Tropical	3
Mediterrâneo	4
Subtropical	5
Temperado	6
Frio	7

Fonte: O autor (2023).

## 5.2 Análise descritiva das variáveis

De acordo com Reis & Reis (2002), a análise descritiva pode ser definida como a utilização da estatística para resumir e descrever as características mais importantes do banco de dados observado da forma mais relevante para possíveis análises.

Ainda segundo Reis & Reis (2002), o termo variável pode ser definido como uma característica que é medida de alguma forma no banco de dados e pode variar seus valores. Essas variáveis podem, ainda, ser classificadas como quantitativas, quando podem ser medidas e apresentam valores numéricos para isso, ou qualitativas, predominantes nesse estudo, quando podem ser separadas por categorias em vez de valores quantitativos.

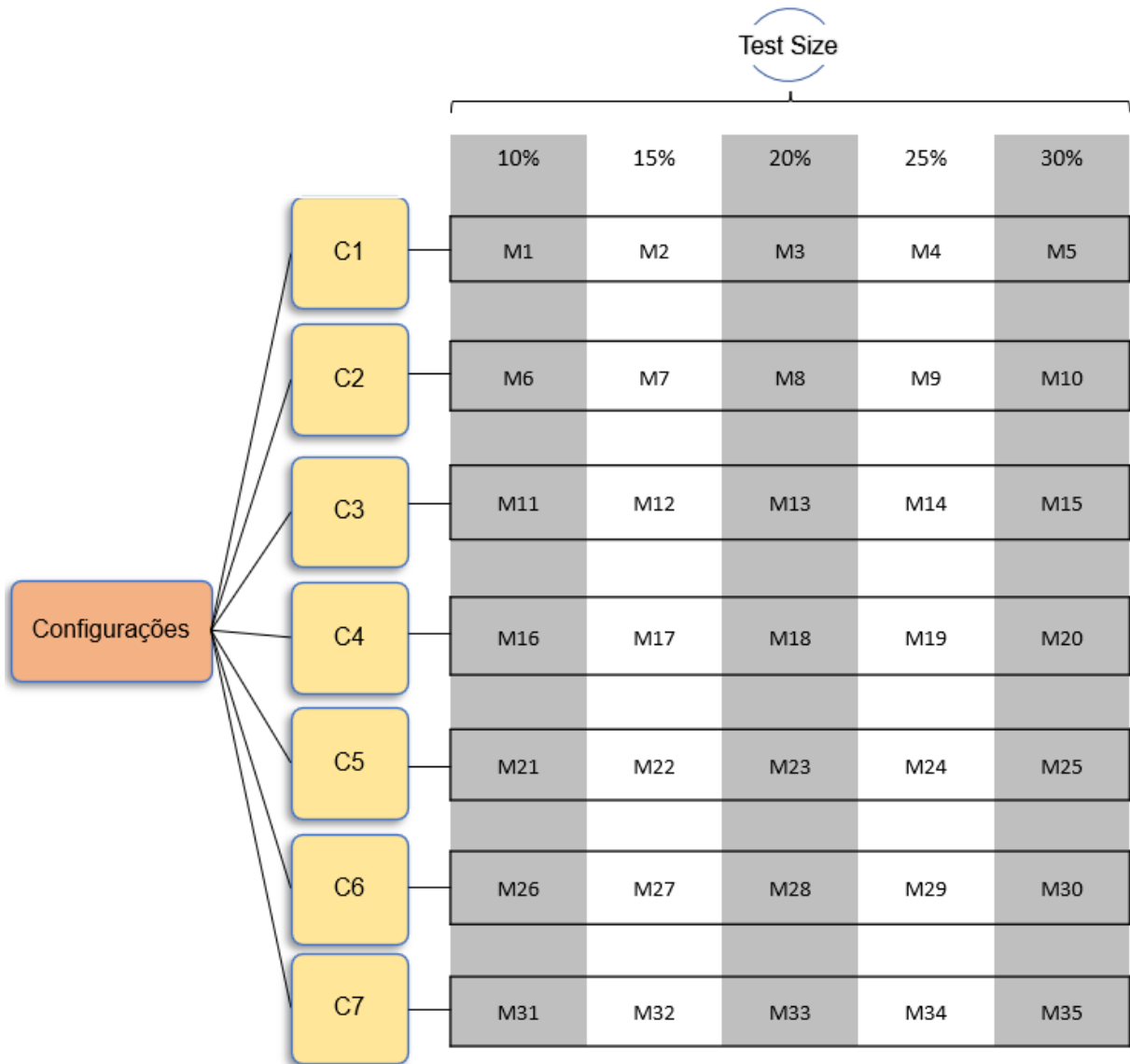
Para os dados utilizados nessa pesquisa, foram gerados gráficos de frequência relativa para as variáveis qualitativas, indicando a porcentagem com que cada categoria aparecia em uma determinada variável no banco de dados, bem como histogramas para as variáveis quantitativas.

### 5.3 Treinamento de modelos preditivos

Considerando diferentes combinações de variáveis com entrada, denominadas de C1 a C7, foram treinados modelos preditivos via KNN, com utilização da linguagem de programação *Python*. Os códigos e ferramentas computacionais utilizados encontram-se no Apêndice B. Todos foram configurados para utilizar a distância euclidiana, com o número de k vizinhos variando de 0 a 35 e estado aleatório igual a zero. O estado aleatório controla o embaralhamento aplicado aos dados antes de aplicar a divisão.

Para cada uma das sete diferentes combinações de variáveis para a entrada foram treinados cinco modelos, com tamanho de amostra de teste/treino igual a, respectivamente, 10%/90%, 15%/85%, 20%/80%, 25%/75% e 30%/70%, gerando no total trinta e cinco modelos preditivos, nomeados de M1 a M35, como apresentado na Figura 18.

Figura 18 - Modelos preditivos gerados para cada configuração.



Fonte: O autor (2023).

As diferentes configurações utilizadas no treinamento dos modelos preditores são apresentadas na tabela 8.



Tabela 8 - Configurações adotadas para treinamento e geração de modelos preditivos.

Denominação da configuração	Variáveis utilizadas como entrada	Variáveis excluídas
C1	Minério que gerou o rejeito armazenado, material de construção da barragem, método construtivo, altura, volume, modo de falha, risco sísmico e clima.	-
C2	Material de construção da barragem, método construtivo, altura, volume, modo de falha, risco sísmico e clima.	Minério que gerou o rejeito armazenado.
C3	Minério que gerou o rejeito armazenado, material de construção da barragem, método construtivo, altura, modo de falha, risco sísmico e clima.	Volume.
C4	Minério que gerou o rejeito armazenado, material de construção da barragem, método construtivo, altura, volume, modo de falha, e clima.	Risco sísmico.
C5	Minério que gerou o rejeito armazenado, material de construção da barragem, método construtivo, altura, volume, modo de falha e risco sísmico.	Clima.
C6	Minério que gerou o rejeito armazenado, material de construção da barragem,	Volume e risco sísmico.

	método construtivo, altura, modo de falha, e clima.	
C7	Minério que gerou o rejeito armazenado, material de construção da barragem, método construtivo, altura, volume e modo de falha.	Risco sísmico e clima.

Fonte: O autor (2023).

### 5.3.1 Pacotes *Python* utilizados

Como pode-se observar no início das informações apresentadas no Apêndice B, os pacotes *Python* utilizados para realização do trabalho foram: *Pandas*, *NumPy* e *Matplotlib*. Além das bibliotecas *Seaborn* e *Scikit-learn*.

O *Pandas*, segundo a plataforma de educação digital Awari (2020), é um dos pacotes mais populares e amplamente utilizados na análise de dados em *Python*. O *NumPy* fornece suporte para *arrays* multidimensionais com alta performance e funções matemáticas avançadas. É amplamente utilizado para operações numéricas, cálculos estatísticos e álgebra linear, sendo a base de muitos outros pacotes de análise de dados em *Python*.

Ainda segundo Awari (2020) o *Matplotlib* é um pacote utilizado para criação e visualização de gráficos. Com ele, é possível criar uma ampla variedade de gráficos, como os de linha, barras, dispersão, histogramas, entre outros.

Também de acordo com Awari (2020) a biblioteca *Seaborn* é utilizada para visualização e plotagem de gráficos estatísticos e a *Scikit-learn* fornece uma ampla variedade de algoritmos e ferramentas para tarefas de aprendizado supervisionado e não supervisionado, como classificação, regressão, agrupamento e pré-processamento de dados. Além de simplificar o processo de desenvolvimento e avaliação de modelos de aprendizado de máquina

#### **5.4 Avaliação e validação dos modelos preditivos gerados**

Com a geração dos modelos foram obtidas, para cada um, as métricas de acurácia e erro. Dessa forma, foi possível avaliar o desempenho de cada modelo na classificação das barragens de rejeito quanto ao seu modo de falha. Além de permitir a análise de como se dava a variação de tais métricas à medida que o valor de K variava.

Para todos os modelos gerados foi realizado, também via *Python*, o teste de otimização do valor de K, considerando o intervalo de zero a trinta e cinco. Nesse teste pode-se observar o erro associado a cada um dos valores de K do intervalo utilizado. Foram selecionados os modelos que apresentaram menor erro de predição (maior acurácia) para cada configuração de C1 a C7.

#### **5.5 Seleção do modelo preditivo**

Em posse das informações referentes ao erro e à acurácia de todos os modelos gerados, do M1 ao M35, bem como dos resultados dos testes de otimização dos valores de K, foi selecionado aquele que apresentou o melhor desempenho. Nesse trabalho, a métrica de acurácia foi a utilizada para avaliar esse desempenho.

#### **5.6 Aplicação do modelo preditivo selecionado**

Após ter avaliado, validado e selecionado o modelo preditivo de melhor desempenho, com base nos testes realizados e métricas avaliadas, pode-se chegar a uma aplicação prática para ele.

Foram selecionadas, através do site <https://app.anm.gov.br/SIGBM/Publico/ClassificacaoNacionalDaBarragem>, todas as barragens classificadas com risco de emergência II e III no Brasil, totalizando um total de dez barragens de rejeito. Todas localizadas no estado de Minas Gerais.

Com a utilização do modelo preditor selecionado, foi possível gerar o resultado de predição para o modo de falha crível para cada uma das dez barragens de rejeito

analisadas. A tabela 9 apresenta quais as barragens selecionadas. Informações quanto às variáveis de cada barragem (dados de entrada) podem ser consultadas no Apêndice C.

Tabela 9 - Barragens de rejeito com nível de emergência 2 e 3 no Brasil.

Barragem de Mineração	Empreendedor	Nível de Alerta e Emergência
Forquilha III	VALE S.A.	3
Dique de Pedra	VALE S.A.	2
Sul Superior	VALE S.A.	3
Forquilha I	VALE S.A.	2
Forquilha II	VALE S.A.	2
Xingu	VALE S.A.	2
Grupo	VALE S.A.	2
Barragem de Rejeitos	ARCELORMITTAL BRASIL S.A.	3
B3/B4	VALE S.A.	2
BARRAGEM B2 AUXILIAR	MINERIOS NACIONAL S.A.	2

Fonte: ANM (2023).

Quanto aos níveis de emergência citados, tem-se que considerar emergências aquelas situações decorrentes de eventos adversos que afetem a segurança da barragem e possam causar danos à sua integridade estrutural e operacional, à preservação da vida, da saúde, da propriedade e do meio ambiente. Assim, as barragens da citadas na tabela 9 foram classificadas de acordo com a Resolução ANM nº 95/2022, com as seguintes definições, apresentadas na Figura 19.

Figura 19 - Classificação do nível de emergência de uma barragem.

<b>NÍVEIS DE EMERGÊNCIA</b>	
<b>Nível</b>	<b>Descrição</b>
<b>Nível 1</b>	As anomalias não comprometem a segurança da barragem no curto prazo, mas devem ser reparadas, controladas ou monitoradas. (Qualquer anomalia com pontuação 10 no Quadro de Estado de Conservação)
<b>Nível 2</b>	As anomalias representam risco à segurança da barragem no curto prazo, devendo ser tomadas providências imediatas para a eliminação do problema
<b>Nível 3</b>	As anomalias representam risco de ruptura iminente, devendo ser tomadas medidas para prevenção e redução dos danos materiais e humanos decorrentes do colapso da barragem.

Fonte: ANM (2022).

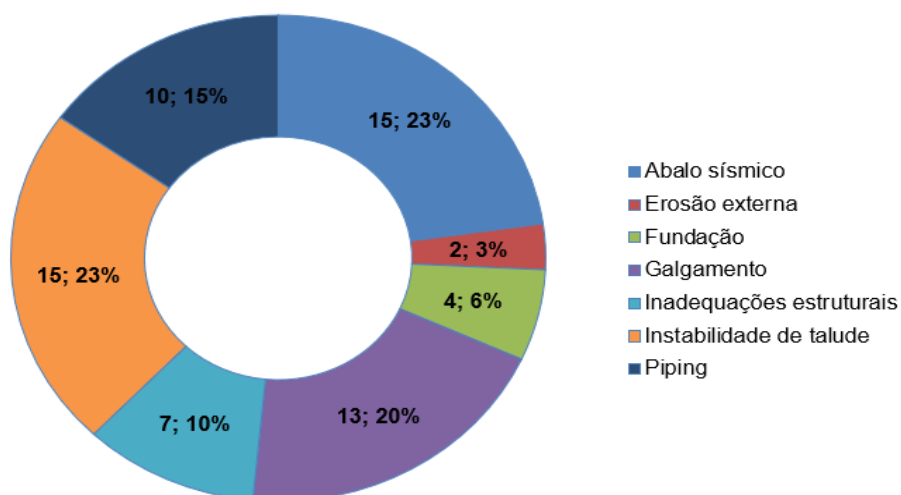
## 6 RESULTADOS

### 6.1 Análise descritiva das variáveis

As Figuras 20 a 22 apresentam, de acordo com o banco de dados utilizado, a frequência relativa de modos de falha das barragens do banco de dados, bem como zona climática, risco sísmico, material de construção da barragem e método construtivo.

Nota-se, na Figura 20, que a maior parte do banco é composto por barragens que romperam por instabilidade de taludes, abalo sísmico ou galgamento, com respectivamente 23%, 23% e 20% do total de ocorrências. Sendo falha por fundação e erosão externa as ocorrências com menor incidência, respectivamente 6% e 3%.

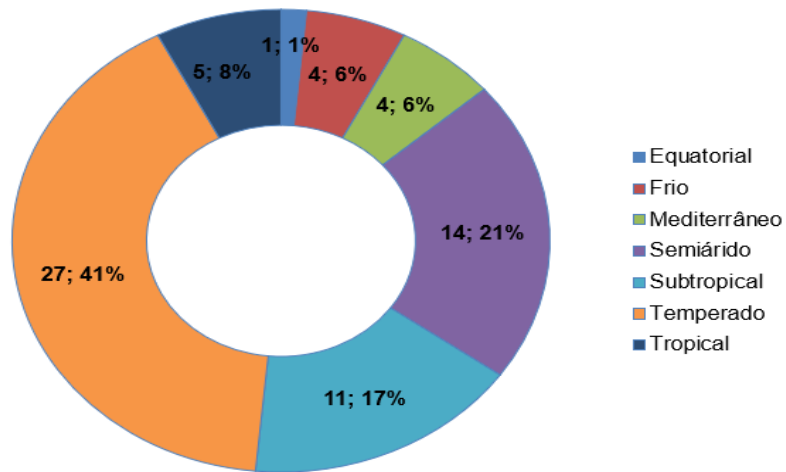
Figura 20 - Quantidade de barragens rompidas por cada modo de falha.



Fonte: O autor (2023).

Já na Figura 21, pode-se observar maior ocorrência de rupturas nos climas temperado e semiárido, com 41% e 21% das ocorrências, respectivamente. E baixo número de ocorrências para os climas tropical, mediterrâneo, frio e equatorial, com 8%, 6%, 6% e 1%, respectivamente.

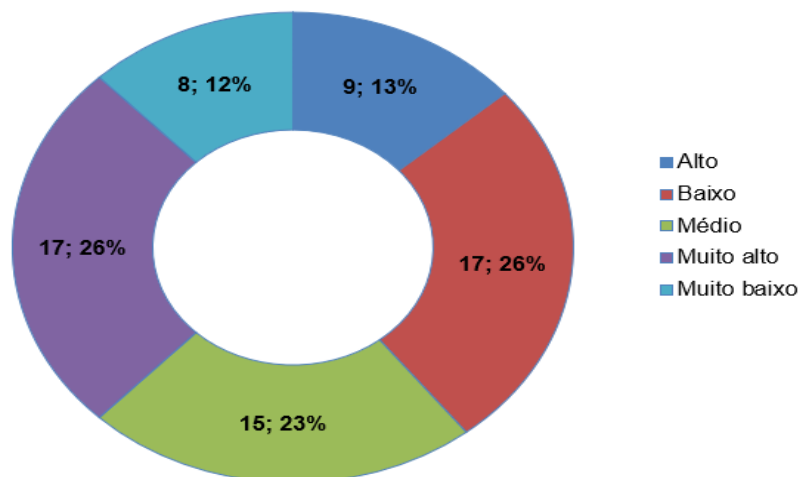
Figura 21 - Quantidade de barragens rompidas por zona de influência climática.



Fonte: O autor (2023).

A Figura 22 apresenta maior ocorrência de falha em locais que apresentam risco sísmico muito alto, baixo e médio, com 26%, 26% e 23% respectivamente. E menor quantidade de ocorrências nas classificações alto e muito baixo, com 13% e 12% respectivamente.

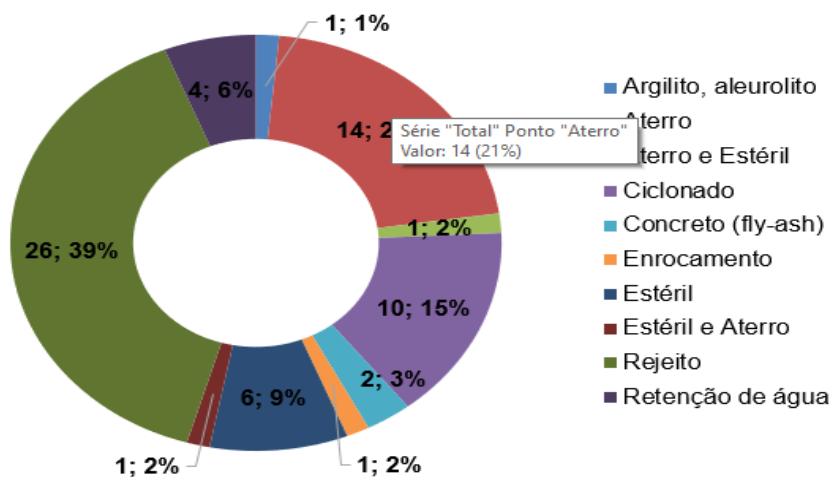
Figura 22 - Quantidade de barragens rompidas por zona de influência sísmica.



Fonte: O autor (2023).

Quanto às informações da Figura 23, observa-se maior ocorrência de ruptura para barragens cujo material de construção foi rejeito, aterro ou material ciclonado, com 39%, 21% e 15% dos casos, respectivamente.

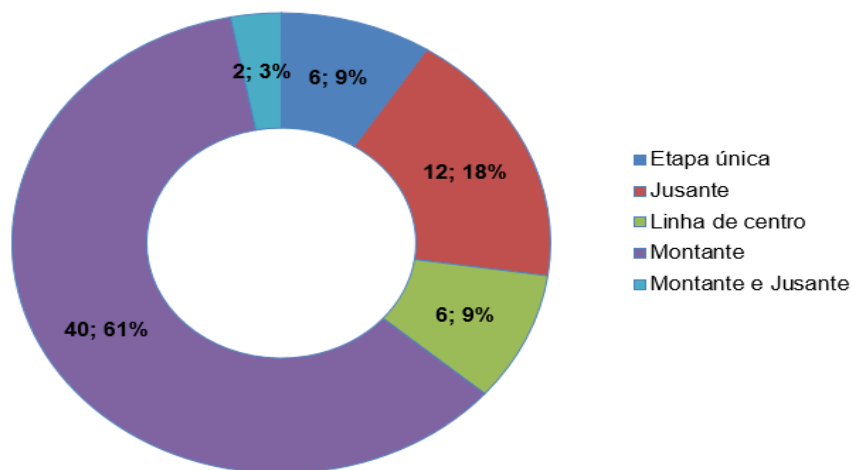
Figura 23 - Quantidade de barragens rompidas de acordo com o material de construção.



Fonte: O autor (2023).

Quanto às rupturas em associação aos métodos construtivos, nota-se na Figura 24 que a maior parte dos casos se dá para barragens construídas pelo método de montante, com 61% dos casos analisados.

Figura 24 - Quantidade de barragens rompidas de acordo com o método de construção.

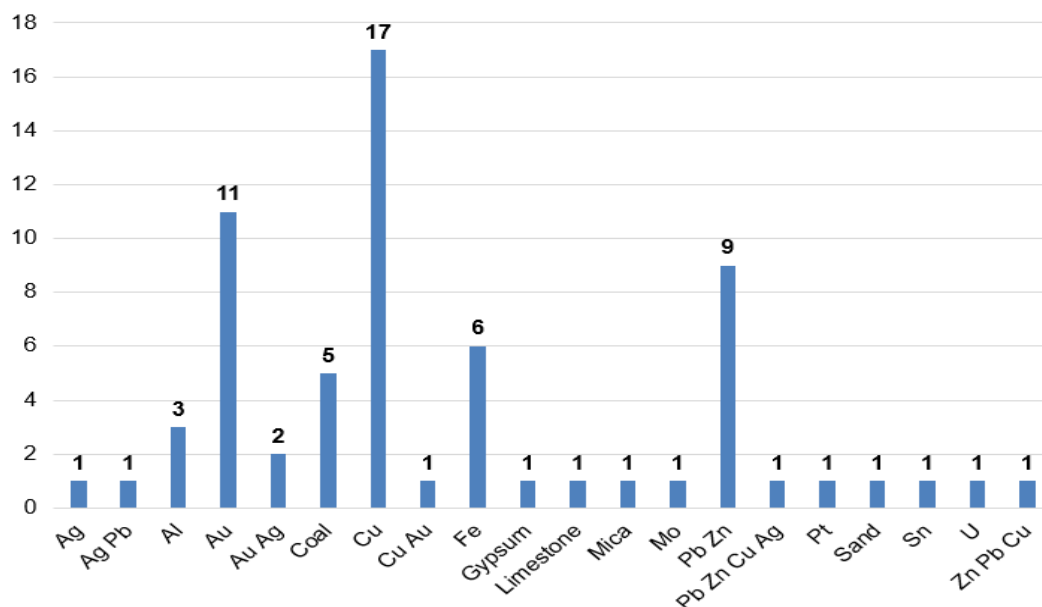


Fonte: O autor (2023)



O gráfico apresentado na Figura 25 indica o número de barragens que possuía ligação com minerações relacionadas a cada tipo de minério indicado na legenda. O maior número de rupturas se dá para barragens relacionadas a minerações de cobre, com 17 das 66 barragens, ouro com 11 barragens e minerações de chumbo e zinco, abarcando 9 barragens rompidas. O maior número de ruptura para barragens relacionadas ao cobre ocorre especialmente devido às características do ambiente onde tais barragens se encontram. No banco de dados utilizado tais barragens se encontravam no Chile, onde há grande instabilidade sísmica.

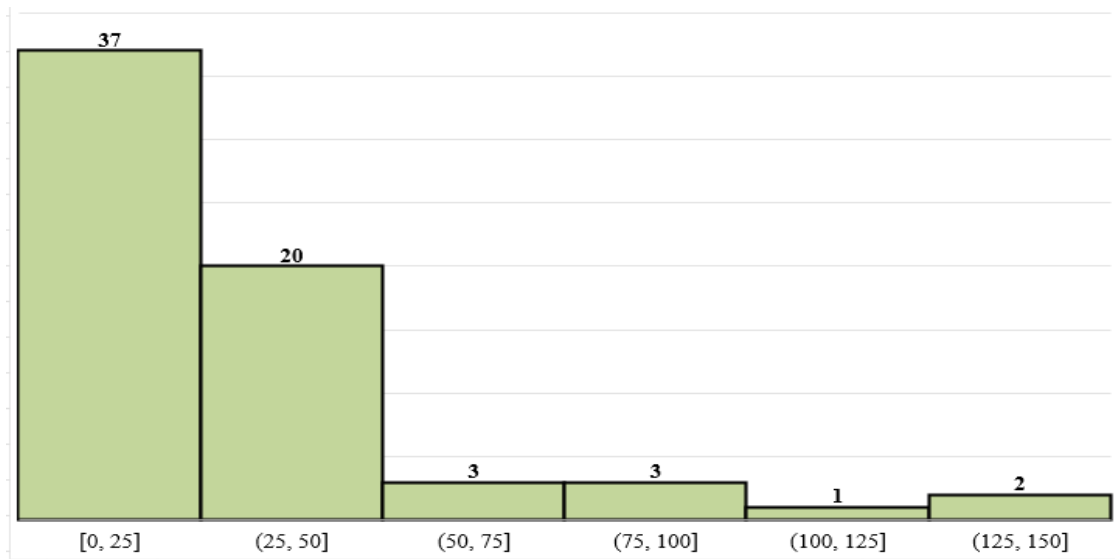
Figura 25 - Quantidade de barragens rompidas relacionadas a cada tipo de mineração.



Fonte: O autor (2023).

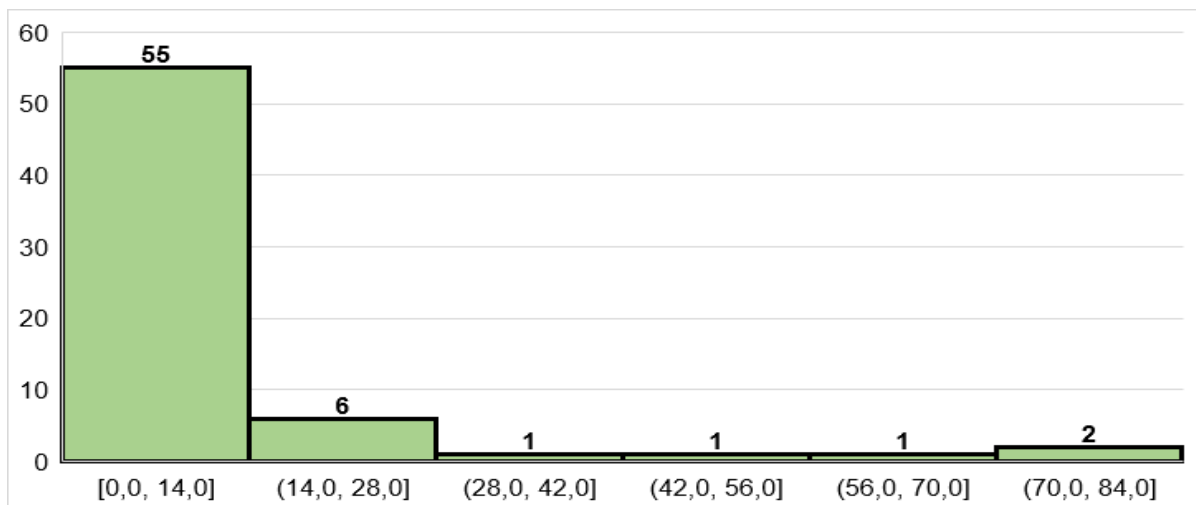
Por fim, foram elaborados os gráficos apresentados nas Figuras 26 e 27, onde tem-se um histograma referente às variáveis altura das barragens e ao volume, respectivamente. Nota-se que a maioria dos casos históricos de barragens que romperam possui altura entre 0 e 50 metros e volumes entre 0 e  $30 \times 10^6$  metros cúbicos.

Figura 26 - Quantidade de barragens por intervalos de altura.



Fonte: O autor (2023).

Figura 27 - Quantidade de barragens por intervalos de volume (106 m3).



Fonte: O autor (2023).

## 6.2 Acurácia dos modelos de predição gerados

Na tabela 10 são apresentados os valores de acurácia obtidos para cada um dos modelos preditivos gerados, bem como os valores ótimos de K, que proporcionaram um menor erro de predição.

Tabela 10 - Resultados dos modelos de predição gerados.

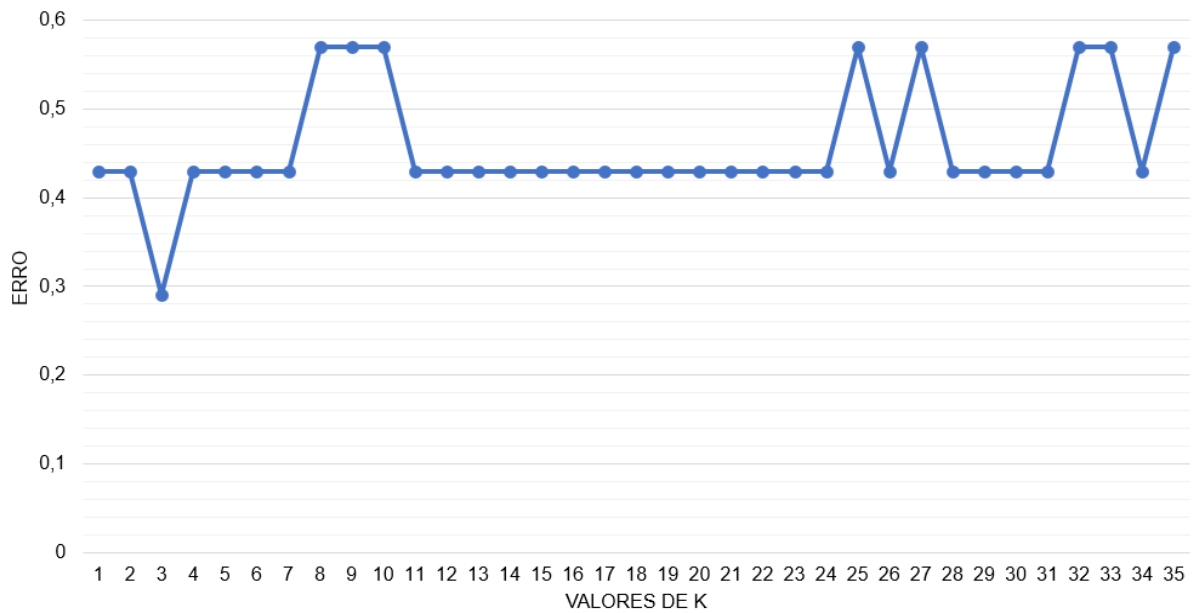
	Test Size	Acurácia	Valor de K		Test Size	Acurácia	Valor de K
C1	0.10	0.57	1, 3, ...	C5	0.10	0.57	1, 8, ...
	0.15	0.50	3, 4, ...		0.15	0.50	24, 25, ...
	0.20	0.36	1 e 9		0.20	0.36	1
	0.25	0.29	1, 3, ...		0.25	0.35	1, 21, ...
	0.30	0.40	4		0.30	0.40	1
C2	0.10	0.57	1, 9, ...	C6	0.10	0.57	3, 6, ...
	0.15	0.40	1, 4, ...		0.15	0.50	8, 9, ...
	0.20	0.29	1, 3, ...		0.20	0.43	9
	0.25	0.35	9		0.25	0.41	8
	0.30	0.40	5		0.30	0.4	7 e 8
C3	<b>0.10</b>	<b>0.71</b>	<b>3</b>	C7	0.10	0.57	24, 25, ...
	0.15	0.50	1, 3, ...		0.15	0.50	25, 26, ...
	0.20	0.43	1		0.20	0.29	7, 21, ...
	0.25	0.35	1		0.25	0.29	6, 7, ...
	0.30	0.45	8		0.30	0.30	5, 6 e 7
C4	0.10	0.57	1, 9, ...				
	0.15	0.50	8, 9, ...				
	0.20	0.43	9				
	0.25	0.41	8				
	0.30	0.40	4, 6 e 7				

Fonte: O autor (2023).

Pode-se observar um melhor desempenho para o banco de dados sem a variável volume, ou seja, a configuração C3, com amostra de teste igual a 10% e K igual a 3, como destacado em amarelo na tabela 10.

Em sequência, o gráfico na Figura 28 apresenta o erro (eixo vertical) associado a cada valor de K (eixo horizontal), para o banco de dados configurado conforme C3, considerando o intervalo de zero a trinta e cinco para o número de vizinhos (K). O valor 1 indica 100%.

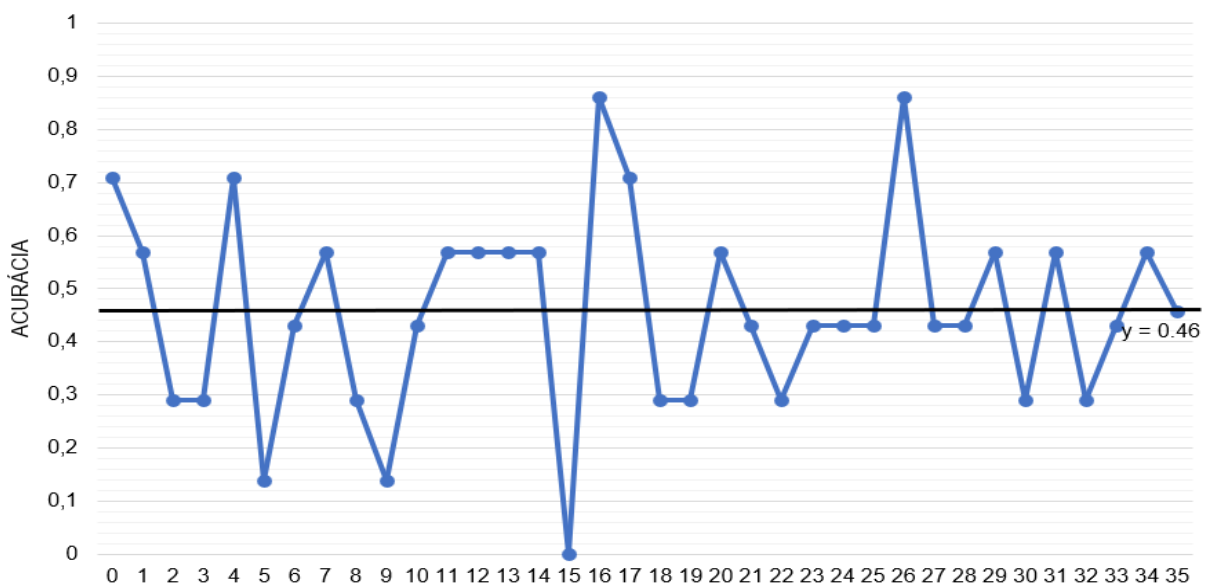
Figura 28 - Erro associado a cada valor de K vizinhos.



Fonte: O autor (2023).

Gerou-se também o gráfico apresentado na Figura 29, onde variou-se o valor do estado aleatório (*random state*), no mesmo intervalo para K, zero a trinta e cinco, a fim de observar a influência exercida na acurácia, obtendo-se um valor médio igual a 46%.

Figura 29 - Variação da acurácia com alteração do estado aleatório.



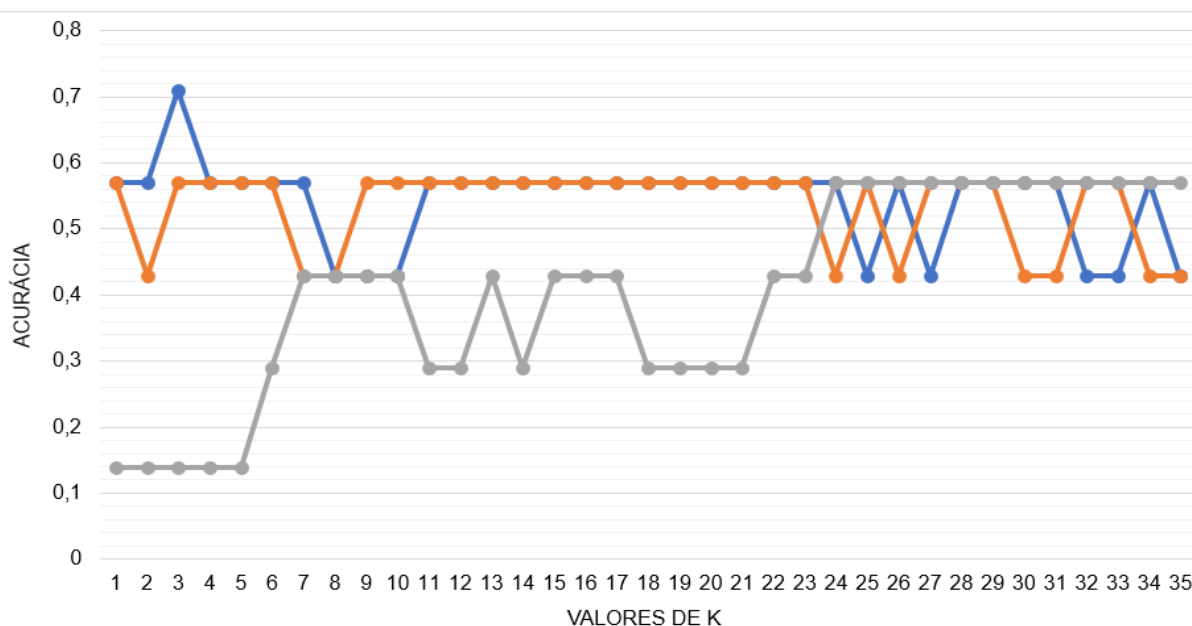
Fonte: O autor (2023).

### 6.3 Análise da influência das variáveis propostas

Tendo em vista os resultados obtidos, foi elaborado o gráfico indicado na Figura 30. Nele é possível observar a variação da acurácia, para o intervalo de valores adotado para K. Foram plotados os dados para a configuração C7 (linha cinza) com objetivo de avaliar a influência que as novas variáveis propostas, risco sísmico e clima, tiveram nos resultados.

Ademais, plotou-se os dados para C1 (linha laranja), banco de dados completo e C3 (linha azul), onde excluiu-se a variável volume. Para tanto, fixou-se o valor de tamanho de amostra de teste em 10%, melhor tamanho de amostra para todos os testes realizados, e estado aleatório igual a zero.

Figura 30 - Variação da acurácia no intervalo de valores de K para C1, C3 e C7.



Fonte: O autor (2023).

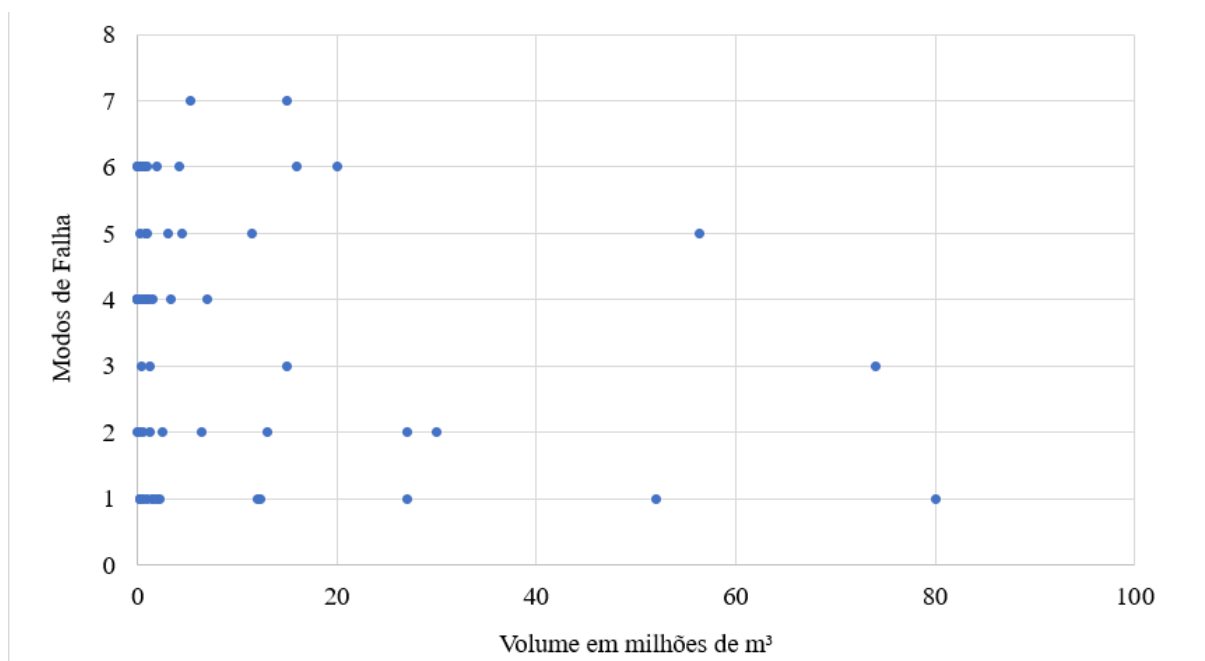
Constatou-se assim, que a inclusão das duas novas variáveis melhorou a acurácia do modelo, como indica a linha laranja em relação à cinza na Figura 30.

### 6.3.1 Justificativa para retirada da variável volume

Para justificar o motivo de a retirada do volume diminuir o erro dos modelos treinados, os gráficos das Figuras 31 a 35 foram elaborados, onde analisa-se a relação do volume com todas as demais variáveis.

Assim, quanto mais dispersos forem os pontos plotados, menor a relação entre elas, o que pode confundir o modelo preditor gerado no momento do treinamento.

Figura 31 - Relação das variáveis volume e modo de falha.

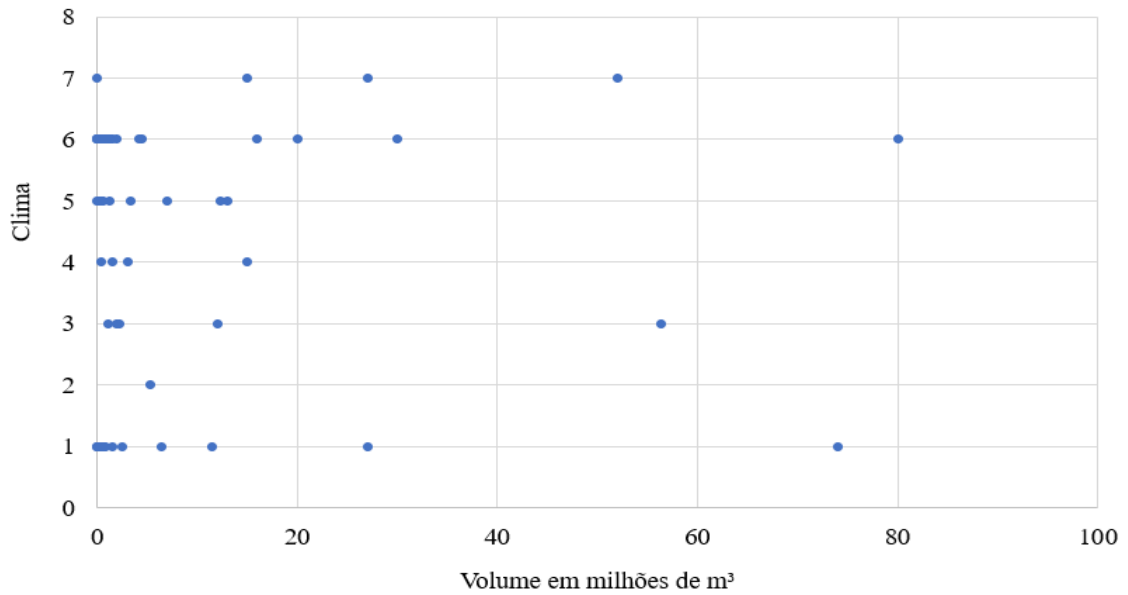


Fonte: O autor (2023).

Pode-se observar maior concentração de todos os modos de ruptura para barragens com volumes inferiores a trinta milhões de metros cúbicos. Como a maior parte das barragens analisadas não ultrapassa tal valor de volume, não se pode definir uma relação específica com os modos de falha.

Nota-se na Figura 32 a mesma situação, considerando agora a variável clima. Não se pode definir uma relação específica entre ela e o volume.

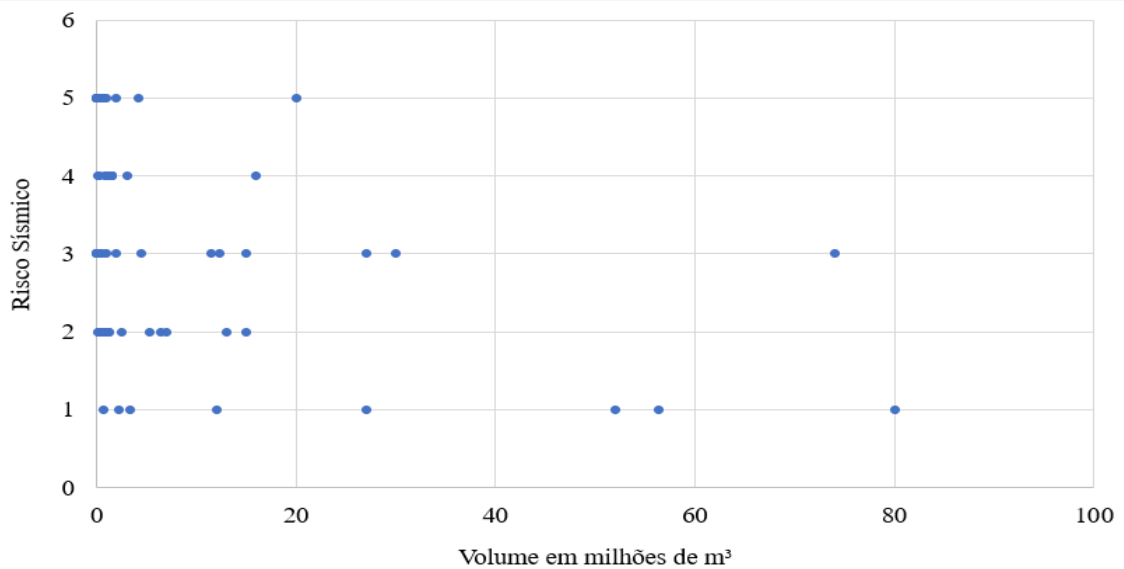
Figura 32 - Relação das variáveis volume e clima.



Fonte: O autor (2023).

A relação entre volume e risco sísmico se observa na Figura 33, trazendo o mesmo aspecto das variáveis analisadas anteriormente.

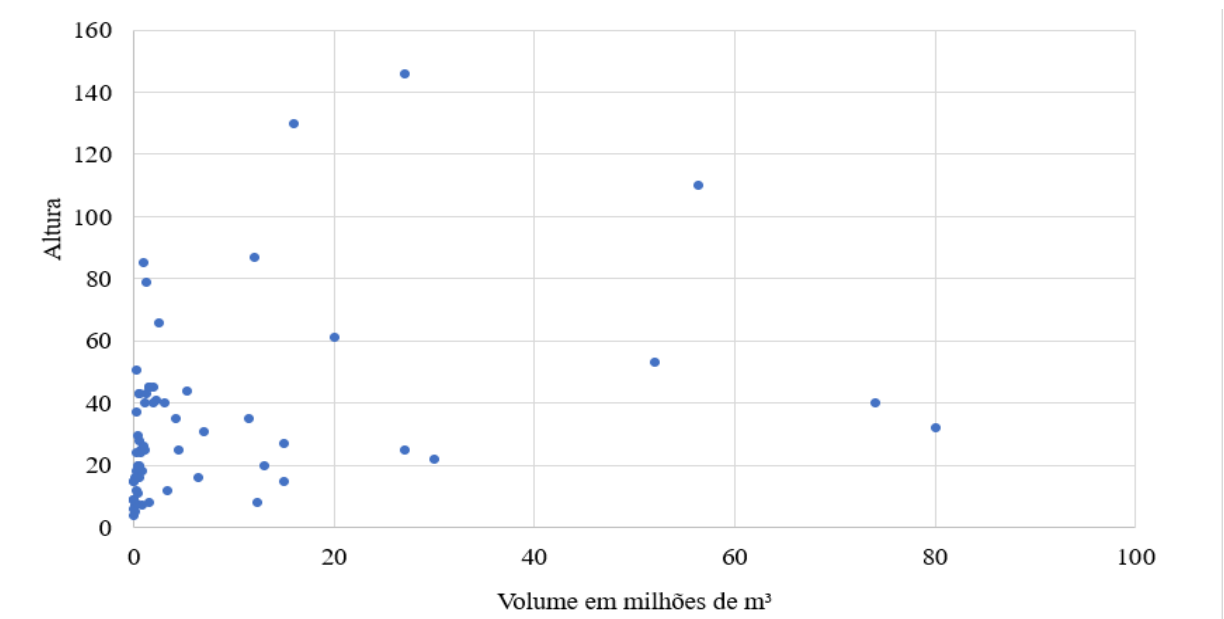
Figura 33 - Relação das variáveis volume e risco sísmico.



Fonte: O autor (2023).

Como esperado, tem-se por meio da Figura 34 a observação de uma relação maior entre volume e altura. Porém, pode-se observar também que não necessariamente se uma barragem possuir um grande volume armazenado sua altura será proporcional, haja vista que as dimensões do reservatório, largura e comprimento, podem aumentar sem aumento de altura.

Figura 34 - Relação das variáveis volume e altura.

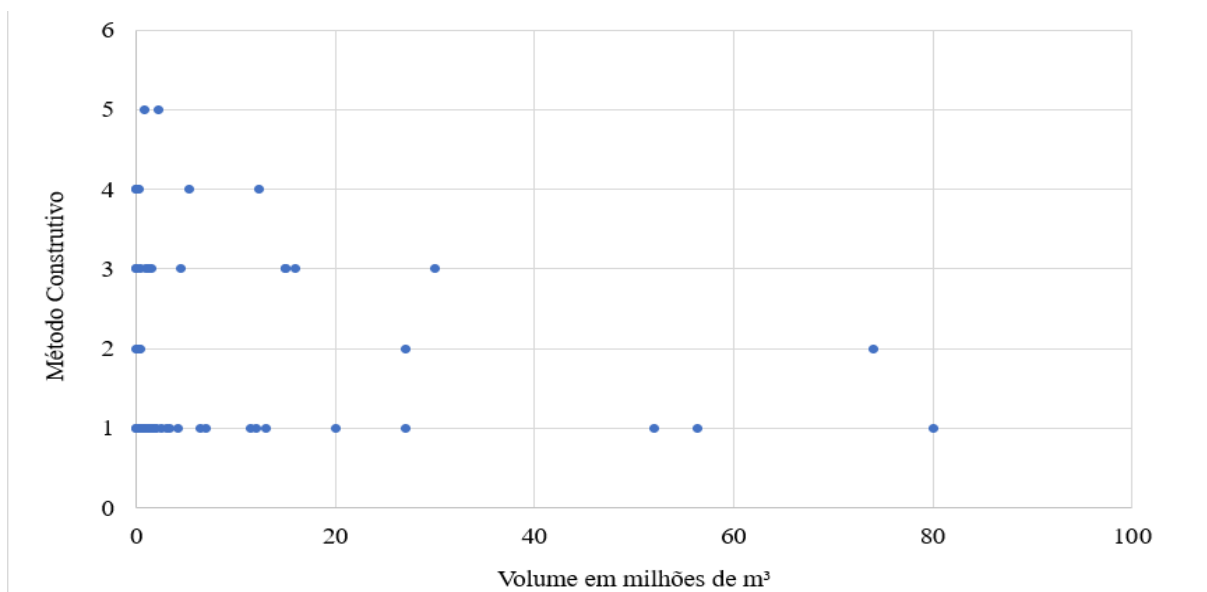


Fonte: O autor (2023).

Quanto à associação ao método construtivo, o mesmo que se notou anteriormente ocorre, baixa relação com o volume do rejeito armazenado na barragem, como observa-se na Figura 35.



Figura 35 - Relação das variáveis volume e método construtivo.



Fonte: O autor (2023).

#### 6.4 Previsão do modo de falha crível das barragens em nível de emergência II e III no Brasil

O modelo preditor de melhor desempenho selecionado, M11, foi utilizado para obter o modo de falha mais provável de ocorrer para cada uma das dez barragens de rejeito classificadas pela ANM com nível de emergência II e III no Brasil, todas localizadas no estado de Minas Gerais.

Na tabela 11 são apresentados os dados da tabela 9, acrescida dos resultados obtidos para o modo de falha crível de cada uma das dez barragens analisadas. No Apêndice B encontra-se a análise computacional que gerou tais resultados.

Tabela 11 - Modo de falha crível para as barragens analisadas.

Barragem de Mineração	Empreendedor	Nível de Alerta e Emergência	Modo de Ruptura Crível
Forquilha III	VALE S.A.	3	Instabilidade de talude (provável liquefação)
Dique de Pedra	VALE S.A.	2	Instabilidade de talude (provável liquefação)

Sul Superior	VALE S.A.	3	Instabilidade de talude (provável liquefação)
Forquilha I	VALE S.A.	2	Instabilidade de talude (provável liquefação)
Forquilha II	VALE S.A.	2	Instabilidade de talude (provável liquefação)
Xingu	VALE S.A.	2	Instabilidade de talude (provável liquefação)
Grupo	VALE S.A.	2	Instabilidade de talude (provável liquefação)
Barragem de Rejeitos	ARCELORMITTAL BRASIL S.A.	3	Instabilidade de talude (provável liquefação)
B3/B4	VALE S.A.	2	Erosão interna ( <i>piping</i> )
BARRAGEM B2 AUXILIAR	MINERIOS NACIONAL S.A.	2	Erosão interna ( <i>piping</i> )

Fonte: O autor (2023).

## 7 CONCLUSÃO

O trabalho desenvolvido evidenciou a importância de se conhecer o modo de falha crível em barragens de rejeito. Este conhecimento pode auxiliar na melhoria da interpretação dos resultados da análise de risco e direcionamento das medidas mitigatória para controle e redução dos riscos de falha. As barragens de rejeito são estruturas que sustentam a indústria da mineração, porém podem ser fontes críticas de perdas de vidas humanas, além de causarem danos ambientais e econômicos consideráveis em caso de falha.

Ademais, é importante não se conhecer apenas os riscos destas estruturas, como também, extrapolar para aplicação mais ampla do conhecimento pode trazer grande benefício para sociedade de maneira geral e ao meio ambiente. Além disso, foi estabelecido e entendido que, a depender do modo de falha crível a análise de estabilidade geotécnica se dará de forma diferente, para se tornar mais efetivo e com resultado mais representativo.

Com base no estudo realizado pode-se perceber também a importância da elaboração de novos métodos e ferramentas, com embasamento estatístico e computacional, que aumentem a eficácia nas ações de combate ao risco de falha em barragens de rejeito. Quando tais métodos puderem ser empregados, nas mais variadas áreas, com a utilização de bancos de dados confiáveis, utilizando de informações e amostras possível, maior será a frequência de testes estatísticos que poderão melhorar a compreensão de cada variável componente dos modelos, trazendo informações que podem ou não contribuir para a interpretação dos resultados das análises.

Neste trabalho, a partir do banco de dados elaborado, foi utilizada a linguagem de programação *Python* para a implementação do método KNN, que variava a porcentagem das amostras de teste do banco, a partir da validação cruzada, a configuração das amostras de treino e teste e o valor de K, ou seja, os vizinhos mais próximos utilizados para a análise. Dessa forma, foram gerados trinta e cinco modelos, classificados de M1 a M35, dos quais os resultados de acurácia permitiram selecionar o de melhor desempenho.

A inclusão das variáveis propostas, risco sísmico e clima, foi positiva no sentido de melhorar o desempenho dos modelos gerados. E no desenvolvimento do estudo a variável volume foi excluída por aumentar o erro das predições.

A partir do uso da técnica de subamostragem aleatória obteve-se uma acurácia média de 46%. O modelo mais preciso dentre os trinta e cinco desenvolvidos foi o M11, no qual obteve-se uma acurácia de 71%. Esse modelo excluiu a variável de volume e foi configurado com k vizinhos igual a 3, tamanho da amostra de teste igual a 10% e estado aleatório igual a zero.

A aplicação desse modelo de KNN foi feita nas dez barragens com nível de emergência II e III no Brasil, todas no estado de Minas Gerais, resultando na predição no modo de falha crível para cada uma delas. Oito foram classificadas com risco de falha por instabilidade de taludes, cuja causa mais provável é a liquefação, sendo elas: Forquilhas I, II e III, Dique de Pedra, Sul Superior, Xingu e Grupo, todas da empresa Vale S.A., e a Barragem de Rejeitos da ArcelorMittal Brasil S.A.

Para as duas barragens restantes foi definida a classificação para risco de falha crível como sendo por erosão interna (*piping*), sendo elas B3/B4 da Vale S.A e Barragem B2 Auxiliar da Minérios Nacional S. A.

Os resultados obtidos apresentam uma primeira análise para o banco de dados utilizado. Medidas como o aumento das informações do banco de dados, com inclusão de outras barragens que sofreram ruptura e a proposição e utilização de novas variáveis nas análises são formas de alcançar a elaboração de modelos preditivos com melhor desempenho. Além da possibilidade de utilização de diferentes distâncias para a técnica de KNN, como a de *Hamming*, *Manhattan*, *Mahalanobis* e *Minkowski*, a fim de comparar os resultados obtidos.

Por fim, outras técnicas mais robustas, como redes neurais artificiais e florestas aleatórias podem ser aplicadas. Construindo então uma mais abrangente rede de análises, com diferentes técnicas estatísticas e computacionais aplicadas a geração de modelos preditores.

## REFERÊNCIAS

Agência Nacional de Mineração - ANM. Veja o que muda com a Resolução ANM nº 95/2022. 2022. Disponível em: <https://www.gov.br/anm/pt-br/assuntos/noticias/veja-o-que-muda-com-a-resolucao-anm-ndeg-95-2022>. Acesso em: 24 ago. 2023.

Agência Nacional de Mineração - ANM. Classificação Nacional de Barragens. Brasil, 2023. Disponível em: <https://app.anm.gov.br/SIGBM/Publico/ClassificacaoNacionalDaBarragem>. Acesso em: 25 jul. 2023.

ANDES – Sindicato Nacional dos Docentes das Instituições de Ensino Superior. 2022. Disponível em: <https://www.andes.org.br/conteudos/noticia/tres-anos-apos-o-rompimento-da-barragem-de-brumadinho-mG-atingidos-ainda-cobram-justica1>. Acesso em: 24 ago. 2023.

ARAUJO, C. B. Contribuição ao estudo do comportamento de barragens de rejeito de mineração de ferro. Tese de Doutorado, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brasil, 2006.

AWARI – Plataforma de Educação Digital. Lista de pacotes *Python* para ciência de dados. 2023. Disponível em: <https://awari.com.br/lista-de-pacotes-python-para-ciencia-de-dados/>. Acesso em: 24 ago. 2023.

AZAM, S.; Li, Q. Tailings dam failures: A review of the last one hundred years. *Geotechnical News*, 28(4), p. 50-54, 2010.

AZANK, F. Modelos de Predição | KNN. Turing Talks, 2019. Disponível em: <https://medium.com/turing-talks/turing-talks-13-modelo-de-predi%C3%A7%C3%A3o-knn-3be880c9b9d1>. Acesso em:

BALBI, D. A. F. Curso Segurança de Barragens. Convênio nº 001/ANA/2011 – SICONV nº 756001/2011. Brasil: ICBA, 2005.

BISHOP, C. M. *Pattern Recognition and Machine Learning*. Springer. Cambridge, 2006.

BRUNTON, S. L.; NOACK, B. R.; KOUMOUTSAKOS, P. Machine Learning for Fluid Mechanics. *Annual Review of Fluid Mechanics*. 2020. v. 52, p. 477-508. Disponível em: <https://doi.org/10.1146/annurev-fluid-010719-060214>. Acesso em: 24 ago. 2023.

CAMPOS, C. A. F. O avanço do Brasil nas políticas de segurança de barragens. Trabalho de Conclusão de Curso. Universidade Federal de Minas Gerais, UFMG, 2022. Disponível em: <https://demin.ufmg.br/downloadtcc.php?f=155>. Acesso em: 26 jul. 2023.

CAPUTO, H. P. *Mecânica dos solos e suas aplicações*. 4. ed. Rio de Janeiro: LTC, v. 3, 1987.

CAPUTO, H. P. *Mecânica dos solos e suas aplicações*. 6. ed. Rio de Janeiro: LTC, v. 2, 2014.

CARDOZO, F. A. C.; PIMENTA, M. M.; ZINGANO, A. C. Métodos Construtivos de Barragens de Rejeitos de Mineração - Uma Revisão. *Holos*, ano 32, v. 8, p. 77-85, 2016. Disponível em: 10.15628/holos.2016.5367. Acesso em: 13 jul. 2023.

CARVALHO, P. S. L. *et al.* Sustentabilidade socioambiental da mineração. BNDES Setorial 47, p. 333-390, 2018.

CASAGRANDE, A. Liquefaction and cyclic deformation of sands - a critical review. 1975, Argentina: 5th Pan-American Conference, p. 81-133, 1975.

CASTRO, G. Liquefaction of sands. 1969. Thesis (Ph.D). Harvard University, Cambridge, Massachusetts, USA, 1969.

CORDEIRO, D. M. Modelo conceitual de transporte integrado à rede de hospitais como atributo de otimização de um sistema urbano de atendimento pré-hospitalar móvel. Tese de Doutorado. Programa de Pós-Graduação em Engenharia Civil. UFPE. Recife, 2012.

COSTA, C. N. Disciplina de fundamentos de geotecnia: Introdução. Faculdade de Ciências e Tecnologia Departamento de Engenharia Civil. Universidade Nova Lisboa, Portugal, 2007. Disponível em: [http://www2.dec.fct.unl.pt/seccoes/S\\_Geotecnia/Fundamentos\\_Geotecnia/2\\_introducao.pdf](http://www2.dec.fct.unl.pt/seccoes/S_Geotecnia/Fundamentos_Geotecnia/2_introducao.pdf). Acesso em: 15 jul. 2023.

DAS, B. M.; SOBHAN, K. Principles of Geotechnical Engineering. 8. ed. Stamford, EUA: CENGAGE Learning, 2014.

DIETERLE, F. Multianalyte Quantifications by Means of Integration of Artificial Neural Networks, Genetic Algorithms and Chemometrics for Time-Resolved Analytical Data. Universidade de Tübingen, Alemanha, 2003.

DYMINSKI, A. S. Noções de estabilidade de taludes e contenções. Centro de Estudos de Engenharia Civil. UFPR. Notas de aula. Disponível em:

<http://www.cesec.ufpr.br/docente/andrea/TC019/TC019/Taludes.pdf>. Acesso em: 14 jul. 2023.

DUARTE, A. P. Classificação das barragens de contenção de rejeitos de mineração e de resíduos industriais no estado de Minas Gerais em relação ao potencial de risco. Dissertação de Mestrado, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil, 2008.

DUDA, R. O.; HART, P. E.; STORK, D. G. Pattern Classification. 2. ed. New York, NY: John Wiley & Sons, 2001.

DUNCAN, J. M.; WRIGHT, S. G.; BRANDON, T. L. Slope Strength and Slope Stability. 2. ed. Nova Jersey: John Wiley & Sons, 2014.

FEMA. Evaluation and Monitoring of Seepage and Internal Erosion. Interagency Committee on Dam Safety (ICODS), Federal Emergency Management Agency (FEMA), 2015.

FENG, S. *et al.* Intelligent driving intelligence test for autonomous vehicles with naturalistic and adversarial environment. Nat Commun, v. 12, p.748, 2021. Disponível em: <https://doi.org/10.1038/s41467-021-21007-8>. Acesso em: 01 nov. 2022.

FONTANA, E. Introdução aos Algoritmos de Aprendizagem Supervisionada. Notas de aula. Departamento de Engenharia Química. UFPR, 2020. Disponível em: [https://fontana.paginas.ufsc.br/files/2018/03/apostila\\_ML\\_pt2.pdf](https://fontana.paginas.ufsc.br/files/2018/03/apostila_ML_pt2.pdf). Acesso em: 15 jul. 2023.

FREIRE, L. Sustentabilidade no Ar. Climas do mundo, 2023. Disponível em: <https://sustentabilidadenoar.com.br/climas-no-mundo/>. Acesso em: 03 jul. 2023.



FREIRE NETO, J. P. Estudo da liquefação em rejeitos e aplicação de metodologia de análise de estabilidade. 2009. 184 f. Dissertação (Mestrado em Geotecnia) - Universidade Federal de Ouro Preto. Escola de Minas. Núcleo de Geotecnia, Ouro Preto - MG, Brasil, 2009.

GODBOLE, S.; SARAWAGI, S. Discriminative Methods for Multi-labeled Classification. In: Dai H., Srikant R., Zhang C. Advances in Knowledge Discovery and Data Mining. PAKDD, 2004. Lecture Notes in Computer Science, v. 3056. Springer, Berlin, Heidelberg. Disponível em: [https://doi.org/10.1007/978-3-540-24775-3\\_5](https://doi.org/10.1007/978-3-540-24775-3_5). Acesso em: 04 abr. 2023.

GOMES, A. S. *et al.* 1º Relatório de Progresso". Grupo de Trabalho de Análise de Riscos em Barragens. Comissão Nacional Portuguesa das Grandes Barragens - CNPGB, Lisboa, 13p, 2005.

GUILLÉN, J. L. C. Modelagem Elasto-plástica da Liquefação Dinâmica de Solos. 2008. 246 f. Tese (Doutorado Engenharia Civil), Pontifícia Universidade Católica do Rio de Janeiro - PUC-RJ, Rio de Janeiro - RJ, Brasil, 2008.

HAIR, J. F. *et al.* Análise Multivariada de Dados. Cengage. 8. ed. 2018.

HAN, J.; KAMBER, M.; PEI, J. Data Mining: Concepts and Techniques. Elsevier, 2011.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer. 2. ed. Stanford, Califórnia, 2009.

INTEROP. Exemplos de Machine Learning aplicados ao dia a dia. Brasil, 2019. Disponível em: <https://www.interop.com.br/blog/exemplos-de-machine-learning/>. Acesso em: 05 abr. 2023.

JOHNSON, F.; ILLES, P. A Classification of Dam Failures. *Water Power Dam Constr*, v. 28, p. 12, 1976.

JONATAS, J. L. Rotura de barragens de aterro por galgamento - Ensaios experimentais com aterros homogêneos. Dissertação de mestrado. Departamento de Engenharia Geográfica, Geofísica e Energia. Universidade de Lisboa. Lisboa, 2013.

KOSSOFF, D. *et al.* Mine tailings dams: characteristics, failure, environmental impacts, and remediation. *Applied Geochemistry*, 51, p. 229-245, 2014.

KUMAR, A. Aprendizado de Máquina - Técnicas de Validação. IA, Ciência de dados, Machine Learning, Blockchain. Disponível em: <https://vitalflux.com/machine-learning-validation-techniques-interview-questions/>. Acesso em: 18 ago. 2023.

LADEIRA, J. E. R. Avaliação de Segurança em barragem de terra, sob o cenário de erosão tubular regressiva, por métodos probabilísticos: O Caso UHE - São Simão. Universidade Federal de Minas Gerais - UFMG. Belo Horizonte, 2007.

LEAO, S. R.; SANTIAGO, A. M. S. Cenário das barragens de rejeito: conhecer para evitar novas catástrofes. *Ambiente & Sociedade*. São Paulo, v. 25, 2022. Disponível em: <https://doi.org/10.1590/1809-4422asoc20210066r1vu2022L2AO>. Acesso em: 10 jul. 2023.

LEI, B. *et al.* Classification, Parameter Estimation and State Estimation: An Engineering Approach Using MATLAB. John Wiley & Sons, 2017.

LOAYZA, F. H. Modelagem do Comportamento Pós-Sismo de uma Barragem de Rejeito. 2009. 220 f. Dissertação (Mestrado em Engenharia Civil), Pontifícia Universidade Católica do Rio de Janeiro - PUC-RJ, Rio de Janeiro - RJ, Brasil, 2009.

LUDERMIR, T. B. Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências. Estudos Avançados, 2021. Disponível em: <https://doi.org/10.1590/s0103-4014.2021.35101.007>. Acesso em: 02 nov. 2022.

MAIORANO, A. C. Modelo de regressão logística. Notas de aula. Instituto de Ciências Matemáticas e de Computação. Universidade de São Paulo - USP, 2021. Disponível em: [https://edisciplinas.usp.br/pluginfile.php/4423145/mod\\_resource/content/1/Regre\\_ssa%CC%83o%20Logistica%20-%20Alexandre%20Maiorano.pdf](https://edisciplinas.usp.br/pluginfile.php/4423145/mod_resource/content/1/Regre_ssa%CC%83o%20Logistica%20-%20Alexandre%20Maiorano.pdf). Acesso em: 15 jul. 2023.

MARIANO, D. C. B. Métricas de avaliação em machine learning: acurácia, sensibilidade, precisão, especificidade, f-score e curva ROC. Brasil, 2021. Disponível em: <https://diegomariano.com/metricas-de-avaliacao-em-machine-learning/>. Acesso em: 30 jun. 2023.

MARQUES FILHO, P. L.; GERALDO, A. Barragens e Reservatórios. In: A. M. dos S. Oliveira & S. N. A. de Brito (Ed.). Geologia de Engenharia. São Paulo: ABGE, 1998.

MARTIN, T. E.; McROBERTS, E. C. Some considerations in the stability analysis of upstream tailings dams. In Proceedings of the Sixth International Conference on Tailings and Mine Waste, v. 99, p. 287-302, 1999. Disponível em: [https://www.researchgate.net/publication/255509946\\_Some\\_considerations\\_in\\_the\\_stability\\_analysis\\_of\\_upstream\\_tailings\\_dams](https://www.researchgate.net/publication/255509946_Some_considerations_in_the_stability_analysis_of_upstream_tailings_dams). Acesso em: 14 jul. 2023.

MATURANO RAFAEL, H. M. A. Análise do Potencial de Liquefação de uma Barragem de Rejeito. 2012. 103 f. Dissertação (Mestrado em Engenharia Civil). Pontifícia Universidade Católica do Rio de Janeiro - PUC-RJ, Rio de Janeiro - RJ, Brasil, 2012.

MITCHELL, T. M. Machine Learning. McGraw-Hill, 1997.

NOGUEIRA, M.; SPRING, J.; PLUMB, C. Vale misled public on dangerous dams, prompting Brazil probe. Reuters Graphics, 2019. Disponível em:

<https://graphics.reuters.com/VALE-DISASTER/0100B29R1CC/index.html>. Acesso em: 12 jul. 2023.

OLIVEIRA, A. R. Comparação de algoritmos de aprendizagem de máquina para construção de modelos preditivos de diabetes não diagnosticado. Dissertação para requisito parcial para Mestrado em Ciência da Computação na Universidade Federal do Rio Grande do Sul. Porto Alegre, 2016.

OTTER, D. W. *et al.* A survey of the usages of deep learning for natural language processing. IEEE Transactions on Neural Networks and Learning Systems, v. 32, p. 604-24, 2020. Disponível em: <https://doi.org/DOI:10.1109/tnnls.2020.2979670>. Acesso em: 10 nov. 2022.

PASSOS, N. C. S. T. Barragem de Rejeito: Avaliação dos Parâmetros Geotécnicos de Rejeito de Minério de Ferro utilizando Ensaio de Campos – Um Estudo de Caso. Trabalho de Conclusão de Curso. UFPR. Brasil, 2009.

PATRO, R. Validação cruzada: K Fold vs Monte Carlo. Escolhendo a técnica de validação correta. Rumo à ciência de dados. Brasil, 2021. Disponível em: <https://towardsdatascience.com/cross-validation-k-fold-vs-monte-carlo-e54df2fc179b>. Acesso em: 25 abr. 2023.

PEREIRA, E. L. Estudo do potencial de liquefação de rejeitos de minério de ferro sob carregamento estático. 2005. 185 f. Dissertação (Mestrado em Engenharia Civil) - Escola de Minas, Universidade Federal de Ouro Preto, Ouro Preto, 2005.

Plano Nacional de Mineração - PNM. Geologia, mineração e transformação mineral. Ministério de Minas e Energia, Brasília, 2011. Disponível em:

[https://ibram.org.br/wp-content/uploads/2021/02/pnm\\_2030.pdf](https://ibram.org.br/wp-content/uploads/2021/02/pnm_2030.pdf). Acesso em: 11 jul. 2023.

POULOS, S. J. The steady state of deformation. Journal of the Geotechnical Engineering Division, ASCE. v. 107, n. GT 5, p. 553-562, 1981.

PRADO, B. Q. M. Análise de Agrupamentos das taxas de incidência de dengue nos estados brasileiros. Trabalho de Conclusão de Curso. Universidade Federal de Uberlândia. Uberlândia, 2015.

RAFAEL, H. M. A. M.; ROMANEL, C. Potencial de Liquefação Estática no Alçamento de uma Barragem de Rejeito. Anais do Congresso Brasileiro de Mecânica de Solos e Engenharia Geotécnica, Goiânia, GO, Brasil, 2014.

RAPOSO, L. K-Nearest Neighbors. Notas de aula. Departamento de Métodos Quantitativos. UNIRIO, 2021. Disponível em: <https://leticiaraposo.netlify.app/>. Acesso em: 09 jun. 2023

REIS, E. A.; REIS, I. A. Análise Descritiva dos dados. Relatório Técnico do Departamento de Estatística da UFMG, 2002. Disponível em: [www.est.ufmg.br](http://www.est.ufmg.br). Acesso em: 11 jul. 2023.

ROLIM, G. S. *et al.* Classificação climática de Köppen e de Thornthwaite e sua aplicabilidade na determinação de zonas agroclimáticas para o estado de São Paulo. *Bragantia*, v. 66, p. 711-720, 2007.

SEDAC - Socioeconomic Data and Applications Center. The areas of the world with the greatest seismic activity, 2014. Disponível em: [sedac.ciesin.columbia.edu](http://sedac.ciesin.columbia.edu). Acesso em: 02 jul. 2022.

SEED, H. B. Considerations in the earthquake-resistant design of earth and rockfill dam. *Géotechnique*, v. 29, n. 3, p. 215-263, 1979.

SOUZA, W. A.; LOTUFO, R. A.; RITTNER, L. Análise comportamental da Optimum Path Forest em diferentes funções métricas. Trabalho de Conclusão de Curso. Universidade Estadual de Campinas. Campinas, 2012.

SOUZA JUNIOR, T.F.; MOREIRA, E. B.; HEINECK, K.S. Barragens de contenção de rejeitos de mineração no Brasil. *Holos*, ano 34, v. 05, 2018. Disponível em: <https://doi.org/10.15628/holos.2018.7423>. Acesso em: 10 jul. 2023.

TERZAGHI, K.; PECK, R. B.; MESRI, G. *Soil Mechanics in Engineering Practice*. 3 ed. New York, USA: John Wiley & Sons, 1996.

THEODORIDIS, S.; KOUTROUMBAS, K. *Pattern Recognition*. Amsterdam: Elsevier, 2009.

TORFI, A. *et al.* Natural language processing advancements by deep learning: A survey. Cornell University, 2020. Disponível em: <https://doi.org/10.48550/arXiv.2003.01200>. Acesso em: 10 nov. 2022.

USACE. Best Practices in Dam And Levee Safety Risk Analysis. U.S. Department of the Interior, Bureau of Reclamation e U.S. Army Corps of Engineers, 2015.

WU, X. *et al.* Top 10 algorithms in data mining. Knowledge and Information Systems, p. 1-37, 2007. Disponível em:

<http://www.ar.sanken.osaka-u.ac.jp/~motoda/papers/10Algorithms.pdf>. Acesso em: 09 mai. 2023.

## APÊNDICES



**Apêndice A – Banco de dados codificado utilizado na geração de modelos preditores.**

Tipo de minério	Método construtivo	Material da barragem	Volume (10 <sup>6</sup> m <sup>3</sup> )	Altura (m)	Risco Sísmico	Clima	Modo de ruptura
1	1	1	12,00	87	1	3	1
2	1	3	2,00	45	3	3	1
1	1	7	0,29	50,7	4	5	1
6	1	1	1,52	45	4	4	1
4	5	2	2,25	41	1	3	1
9	1	10	52,00	53	1	7	1
9	1	4	0,30	37	2	6	1
1	1	2	0,40	29,5	5	4	1
6	4	1	12,30	8	3	5	1
1	1	9	80,00	32	1	6	1
4	3	4	1,54	8	4	1	1
1	1	2	27,00	25	1	7	1
6	1	1	1,00	25	3	6	1
17	1	2	0,30	18	3	5	1
3	1	4	0,50	43	3	1	1
2	3	6	30,00	22	3	6	2
3	1	4	0,50	43	2	6	2
4	2	4	0,12	7	2	6	2
8	2	2	27,00	146	3	1	2
4	4	4	0,12	5	4	1	2
3	1	2	2,50	66	2	1	2
12	1	1	13,00	20	2	5	2
6	3	4	1,23	43	2	5	2
19	1	1	6,36	16	2	1	2
20	4	1	0,027	9	5	7	2
5	2	3	74,00	40	3	1	3
6	3	7	15,00	27	3	4	3
9	3	3	1,23	79	2	6	3
11	2	4	0,37	11	2	1	3
4	2	4	0,025	4	5	6	4
4	1	1	7,04	31	2	5	4
14	3	4	0,037	9	3	1	4
16	1	4	3,30	12	1	5	4
10	4	4	0,20	7,5	2	6	4
3	1	1	1,10	40	4	3	4

15	4	4	0,038	6	3	5	4
4	1	1	0,680	25	1	1	4
6	1	1	0,75	18	2	1	4
18	3	4	0,037	9	3	1	4
13	1	3	1,08	25	2	6	4
9	1	7	0,50	16	2	6	4
6	1	1	1,52	45	4	6	4
1	1	3	56,40	110	1	3	5
4	5	1	0,80	7	4	6	5
2	3	6	4,50	25	3	6	5
9	3	3	1,00	85	2	6	5
7	2	1	0,215	24	3	1	5
6	1	1	3,00	40	4	4	5
14	1	1	11,48	35	3	1	5
3	1	2	0,08	16	5	6	6
3	3	7	16,00	130	4	6	6
3	1	2	2,00	40	5	6	6
3	1	1	0,70	24	5	6	6
4	1	1	0,48	19	5	5	6
4	1	1	0,48	28	5	5	6
6	1	1	0,30	12	5	5	6
3	1	1	4,25	35	5	6	6
3	3	2	0,35	19	5	6	6
3	1	1	0,985	26	5	6	6
3	1	1	0,043	15	5	6	6
3	1	1	0,45	20	5	6	6
3	1	1	0,03	15	5	6	6
3	1	1	0,50	20	5	6	6
3	1	2	20,00	61	5	6	6
3	3	8	15,00	15	2	7	7
4	4	5	5,25	44	2	2	7

Fonte: O autor (2023).

## Apêndice B – códigos e ferramentas computacionais utilizados na elaboração dos modelos preditores e aplicação em banco de dados para geração de resultados.

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
%matplotlib inline
```

```
df = pd.read_excel('Banco de dadosSemVolume.xlsx')
```

```
df.head()
```

	Tipo de minério	Método construtivo	Material da barragem	Altura	Risco Sísmico	Climas	Modo de ruptura
0	1	1	1	87.0	1	3	1
1	2	1	3	45.0	3	3	1
2	1	1	7	50.7	4	5	1
3	6	1	1	45.0	4	4	1
4	4	5	2	41.0	1	3	1

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 66 entries, 0 to 65
Data columns (total 7 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Tipo de minério       66 non-null    int64
1   Método construtivo    66 non-null    int64
2   Material da barragem  66 non-null    int64
3   Altura                66 non-null    float64
4   Risco Sísmico         66 non-null    int64
5   Climas                66 non-null    int64
6   Modo de ruptura       66 non-null    int64
dtypes: float64(1), int64(6)
memory usage: 3.7 KB
```

```
df.columns
```

```
Index(['Tipo de minério', 'Método construtivo', 'Material da barragem',
      'Altura', 'Risco Sísmico', 'Climas ', 'Modo de ruptura'],
      dtype='object')
```

```
from sklearn.preprocessing import StandardScaler
```

```
scaler = StandardScaler()
```

```
scaler.fit(df.drop('Modo de ruptura', axis=1))
```

```
StandardScaler()
```

```
df_normalizado = scaler.transform(df.drop('Modo de ruptura', axis=1))
```

```
df_param = pd.DataFrame(df_normalizado, columns = df.columns[:-1])
```

```
df_param.head()
```

	Tipo de minério	Método construtivo	Material da barragem	Altura	Risco Sísmico	Climas
0	-1.055112	-0.716772	-0.840072	1.912018	-1.566600	-0.693807
1	-0.873291	-0.716772	0.026252	0.418313	-0.110324	-0.693807
2	-1.055112	-0.716772	1.758900	0.621030	0.617814	0.280475
3	-0.146008	-0.716772	-0.840072	0.418313	0.617814	-0.206666
4	-0.509650	2.662294	-0.406910	0.276055	-1.566600	-0.693807

```
from sklearn.model_selection import train_test_split
```

```
X_train, X_test, y_train, y_test = train_test_split(df_param, df['Modo de ruptura'],  
                                                    test_size=0.1, random_state=0)
```

```
from sklearn.neighbors import KNeighborsClassifier
```

```
knn = KNeighborsClassifier(n_neighbors = 1)
```

```
knn.fit(X_train, y_train)
```

```
KNeighborsClassifier(n_neighbors=1)
```

```
pred = knn.predict(X_test)
```

```
from sklearn.metrics import classification_report, confusion_matrix
```

```
print(classification_report(y_test, pred))
```

	precision	recall	f1-score	support
3	0.00	0.00	0.00	1
4	0.50	0.50	0.50	2
5	0.00	0.00	0.00	1
6	0.75	1.00	0.86	3
accuracy			0.57	7
macro avg	0.31	0.38	0.34	7
weighted avg	0.46	0.57	0.51	7

```
print(confusion_matrix(y_test, pred))
```

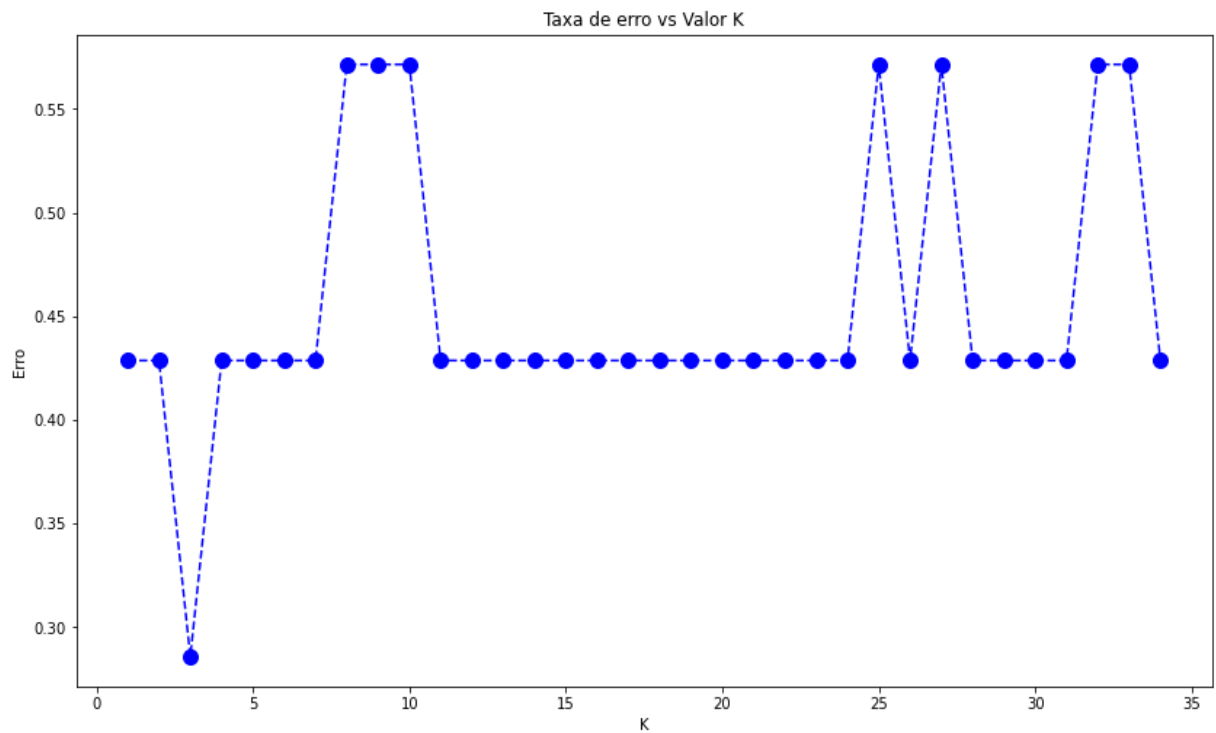
```
[[0 1 0 0]
 [0 1 0 1]
 [1 0 0 0]
 [0 0 0 3]]
```

```
error_rate=[]
```

```
for i in range(1, 35):
    knn = KNeighborsClassifier(n_neighbors=i)
    knn.fit(X_train, y_train)
    pred = knn.predict(X_test)
    error_rate.append(np.mean(pred!=y_test))
```

```
plt.figure(figsize=(14, 8))
plt.plot(range(1, 35), error_rate, color='blue', linestyle='dashed', marker='o', markersize=10)
plt.xlabel('K')
plt.ylabel('Erro')
plt.title('Taxa de erro vs Valor K')
```

```
Text(0.5, 1.0, 'Taxa de erro vs Valor K')
```



```
knn = KNeighborsClassifier(n_neighbors= 3)
knn.fit(X_train, y_train)
pred = knn.predict(X_test)
```

```
print(confusion_matrix(y_test, pred))
print('\n')
print(classification_report(y_test, pred))
```

```
[[0 0 0 0 0]
 [0 1 0 0 0]
 [0 0 1 0 1]
 [1 0 0 0 0]
 [0 0 0 0 3]]
```

	precision	recall	f1-score	support
1	0.00	0.00	0.00	0
3	1.00	1.00	1.00	1
4	1.00	0.50	0.67	2
5	0.00	0.00	0.00	1
6	0.75	1.00	0.86	3
accuracy			0.71	7
macro avg	0.55	0.50	0.50	7
weighted avg	0.75	0.71	0.70	7

```
ma = pd.read_excel('Cópia de Avaliação Barragens.xlsx')
```

```
ma.head()
```

	Tipo de minério	Método construtivo	Material da barragem	Altura	Risco Sísmico	Climas	Modo de ruptura
0	1	1	4	77.00	1	3	NaN
1	1	4	5	19.60	1	3	NaN
2	1	1	9	85.00	1	3	NaN
3	1	1	4	96.83	1	3	NaN
4	1	1	4	90.87	1	3	NaN

```
scaler.fit(ma.drop('Modo de ruptura', axis=1))
```

```
StandardScaler()
```

```
ma_normalizado = scaler.transform(ma.drop('Modo de ruptura', axis=1))
```

```
ma_param = pd.DataFrame(ma_normalizado, columns = ma.columns[:-1])
```

```
ma_param.head()
```

	Tipo de minério	Método construtivo	Material da barragem	Altura	Risco Sísmico	Climas
0	0.0	-0.333333	-0.652328	0.482509	0.0	0.0
1	0.0	3.000000	-0.326164	-1.750862	0.0	0.0
2	0.0	-0.333333	0.978492	0.793781	0.0	0.0
3	0.0	-0.333333	-0.652328	1.254073	0.0	0.0
4	0.0	-0.333333	-0.652328	1.022176	0.0	0.0

```
predicao = knn.predict(ma_param)
```

```
print(predicao)
```

```
[1 1 1 1 1 1 1 1 2 2]
```

**Apêndice C – Banco de dados codificado utilizado na predição de modo de falha para as barragens do estado de minas gerais.**

Tipo de minério	Método construtivo	Material da barragem	Volume (10 <sup>6</sup> m <sup>3</sup> )	Altura (m)	Risco Sísmico	Clima	Modo de ruptura
1	1	4	19,48	77,00	1	3	
1	4	5	0,50	19,60	1	3	
1	1	9	5,94	85,00	1	3	
1	1	4	12,76	96,83	1	3	
1	1	4	22,78	90,87	1	3	
1	1	1	6,17	70,00	1	3	
1	1	4	1,96	41,41	1	3	
1	1	9	5,03	85,00	1	3	
1	1	11	0,90	47,28	1	3	
1	1	9	4,50	33,00	1	3	