



UNIVERSIDADE FEDERAL DE OURO PRETO

ESCOLA DE MINAS

DEPARTAMENTO DE ENGENHARIA DE MINAS



**PROPOSTA DE MODELO DE PREDIÇÃO DA CONDIÇÃO DE ESTABILIDADE DE
TALUDES DE MINA COM USO DO ALGORITMO *K-NEAREST NEIGHBORS***

VITOR MOREIRA CANGUSSU

OURO PRETO

2022

VITOR MOREIRA CANGUSSU

**PROPOSTA DE MODELO DE PREDIÇÃO DA CONDIÇÃO DE ESTABILIDADE DE
TALUDES DE MINA COM USO DO ALGORITMO *K-NEAREST NEIGHBORS***

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

Orientadora: Tatiana Barreto dos Santos

OURO PRETO

2022

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

C222p Cangussu, Vitor Moreira.

Proposta de modelo de predição da condição de estabilidade de taludes de mina com uso do algoritmo k-nearest neighbors. [manuscrito] / Vitor Moreira Cangussu. - 2022.

56 f.: il.: color., gráf., tab..

Orientadora: Profa. Dra. Tatiana Barreto dos Santos.

Monografia (Bacharelado). Universidade Federal de Ouro Preto. Escola de Minas. Graduação em Engenharia de Minas .

1. Taludes (Mecânica do solo) - Estabilidade. 2. Análise de riscos. 3. Aprendizado de máquina. I. Santos, Tatiana Barreto dos. II. Universidade Federal de Ouro Preto. III. Título.

CDU 624.121.537

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



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE OURO PRETO
REITORIA
ESCOLA DE MINAS
DEPARTAMENTO DE ENGENHARIA DE MINAS



FOLHA DE APROVAÇÃO

Vitor Moreira Cangussu

Proposta de Modelo de Predição da Condição de Estabilidade de Taludes de Mina com Uso do Algoritmo K-Nearest Neighbors

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 03 de junho de 2022

Membros da banca

Dr^a. Tatiana Barreto dos Santos - Orientador(a) (UFOP)
Dr. José Margarida da Silva - (UFOP)
Dr. Daniel Silva Jaques - (UFOP)

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 07/06/2022



Documento assinado eletronicamente por **Tatiana Barreto dos Santos, PROFESSOR DE MAGISTERIO SUPERIOR**, em 10/06/2022, às 13:13, 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, informando o código verificador **0342902** e o código CRC **F7AEF29C**.

AGRADECIMENTOS

Agradeço primeiramente aos meus pais, Delvart e Patrícia, e ao meu irmão Dudu pelo apoio incondicional em todos esses anos. Agradeço, também, a todos os meus familiares por todo o suporte dado durante toda a minha trajetória para chegar até aqui, que começou muito antes da graduação.

A todos de Águas Formosas que estão comigo nesse tempo, que sempre me acompanharam e são uma família para mim. Também aos meus colegas e amigos da UFOP, em especial aos do 17.1, por toda a amizade e apoio nesses anos em Ouro Preto.

Às entidades da UFOP que tive a honra de participar e contribuíram muito para o meu conhecimento e crescimento pessoal: a UFOP SEG *Student Chapter*, o *Mining Team* da Escola de Minas, a Liga Acadêmica de Tratamento de Minério e o CREA Jr. - Núcleo Ouro Preto.

A todos os professores que tive durante a graduação, em especial ao professor Vladimir pela oportunidade de desenvolver minha iniciação científica e à professora Tatiana que me orientou durante todo o desenvolvimento deste trabalho.

À TRT Consultoria em Mineração pelo estágio que me abriu portas em um momento difícil e, por fim, agradeço à Hexagon pela oportunidade de ter minha primeira vivência profissional, onde estou tendo uma nova experiência de muito crescimento técnico e pessoal.

RESUMO

A análise de risco de rupturas em taludes rochosos em uma mina a céu aberto pode ser definida como o estudo realizado sobre os riscos de rupturas nestas estruturas e é fundamental para a garantia de segurança do projeto. Para isso, se faz necessário o controle dessas obras, analisando todas as variáveis que podem influenciar em sua estabilidade. Quando esse controle não é executado de maneira adequada, o risco de uma ruptura inesperada de uma obra cresce e, conseqüentemente, pode culminar em ferimento ou até morte de pessoas envolvidas, danos ao meio ambiente ao redor, além do prejuízo financeiro com equipamentos, atividades paralisadas, dentre outros. Este trabalho tem o objetivo de utilizar aprendizado de máquina, através do método KNN (*K-Nearest Neighbor*), para propor um modelo capaz de prever a condição de estabilidade de taludes de mina a partir das suas características. Para obtenção do modelo e predição a condição de estabilidade de taludes de mina foi utilizado o banco de dados de Santos (2019), que é constituído por 88 taludes localizados ao redor do mundo. O modelo obtido, quando testado, apresentou uma acurácia de 83,84%. Por fim, foi feita uma comparação do modelo obtido como modelo proposto por Santos (2019), onde foi utilizada análise discriminante com o mesmo banco de dados e obteve-se uma acurácia de 88,64%.

Palavras-chave: Análise de risco, taludes rochosos, aprendizado de máquina, *K-Nearest Neighbor*.

ABSTRACT

The risk analysis of ruptures in rocky slopes in an open pit mine can be defined as the study carried out on the risks of ruptures in these structures and is fundamental to guarantee the safety of the project. When this control is not carried out properly, the risk of an unexpected slope failure grows and, consequently, can culminate in injury or even death of people involved, damage to the environment, financial losses (equipment loss, paralyzed activities, among others). This work aims to present the application of machine learning, through the KNN (K-Nearest Neighbor) method, to propose a model capable of predicting the stability condition of mine slopes based on their characteristics. To obtain the model and predict the stability condition of mine slopes, Santos (2019) database, which consists of 88 slopes located around the world, was used. The model obtained, when tested, presented an accuracy of 83.84%. Finally, a comparison was made of the model obtained with the model proposed by Santos (2019), where discriminant analysis was used with the same database and an accuracy of 88.64% was obtained.

Keywords: Risk analysis, rocky slopes, machine leaning, K-Nearest Neighbor.

LISTA DE FIGURAS

Figura 1 - Escalas de taludes de mina (Santos, 2019)	14
Figura 2 - Apresentação dos principais parâmetros geomecânicos avaliados na caracterização do maciço rochoso (LACERDA, S.G. et al., 2019)	13
Figura 3 - Medição de espaçamento de descontinuidades em uma face exposta (Filho & Nummer, 2014)	14
Figura 4 - Perfis de rugosidade e seus valores de JRC (Adaptado de Barton & Choubey, 1977).....	15
Figura 5 - Definição da orientação do plano de uma descontinuidade N60E,30SE: (a) vista simétrica; (b) vista em planta (Filho & Nummer, 2014).....	16
Figura 6 – parâmetros para determinação do GSI (Adaptado de Hoek & Marinos, 2000)	19
Figura 7 - Fluxograma de sequência de aprendizagem de máquina (Adaptado de Pham, et al. 2021). 20	
Figura 8 - Modelo gráfico 2D do funcionamento do método KNN (Adaptado de Azank, 2019).....	22
Figura 9 - Fluxograma do trabalho.....	31
Figura 10 - gráfico de quantidade de taludes em cada categoria de rugosidade.....	33
Figura 11 - Gráfico de quantidade de taludes em cada categoria de preenchimento.....	34
Figura 12 - Gráfico de quantidade de taludes em cada categoria de alteração da rocha.	34
Figura 13 - Gráfico de quantidade de taludes em cada categoria de condição de água subterrânea	35
Figura 14 - Gráfico de quantidade de taludes em cada categoria de orientação da descontinuidade principal	35
Figura 15 - Gráfico de quantidade de taludes em cada categoria de método de desmonte.	36
Figura 16 - Gráfico de acurácia das sementes a partir dos valores de K considerando 70% de amostra de treino.....	38
Figura 17 - Valor médio das acurácias para cada valor de K utilizando 70% das amostras de treino. 38	
Figura 18 - Gráfico de acurácia das sementes a partir dos valores de K considerando 80% de amostra de treino.....	40
Figura 19 - Valor médio das acurácias para cada valor de K utilizando 80% das amostras de treino. 40	
Figura 20 - Gráfico de acurácia das sementes a partir dos valores de K considerando 90% de amostra de treino.....	42
Figura 21 - Valor médio das acurácias para cada valor de K utilizando 90% das amostras de treino. 42	

LISTA DE TABELAS

Tabela 1 - Classificação das persistências (Filho e Nummer, 2014).....	14
Tabela 2 - Classificação das aberturas (Adaptado de ISRM, 1978)	16
Tabela 3 - Classes da qualidade de classificação do maciço rochoso de acordo com o RQD (Adaptado de Deere, 1989)	17
Tabela 4 - Classificação de maciços rochosos segundo o RMR (Adaptado de Bieniawski, 1989)	18
Tabela 5 - Classificação do maciço rochoso pelo sistema Q (Adaptado de Bieniawski, 1989).....	18
Tabela 6 - Exemplo de matriz de confusão para três classes (Santos, 2019)	23
Tabela 7 - Minas do banco de dados (Santos, 2019).....	27
Tabela 8 - Valores associados aos parâmetros (Santos, 2019)	28
Tabela 9 - Taludes do banco de dados (Santos, 2019)	29
Tabela 10 - Análise descritiva das variáveis quantitativas do banco de dados.....	32
Tabela 11 - Acurácias do método variando as sementes e o K, considerando 70% de amostra de treino.	37
Tabela 12 - Acurácias do método variando as sementes e o K, considerando 80% de amostras de treino	39
Tabela 13 - Acurácias do método variando as sementes e o K, considerando 90%.....	41
Tabela 14 - Banco de dados utilizado no trabalho	52

SUMÁRIO

1. INTRODUÇÃO.....	8
2. OBJETIVOS	10
2.1 Objetivo geral	10
2.2 Objetivos específicos.....	10
3. REVISÃO BIBLIOGRÁFICA.....	11
3.1. Conceitos iniciais em estabilidade de taludes de cava.....	11
3.2. Maciços rochosos.....	12
3.3. Aprendizado de máquina.....	19
3.3.1 Aprendizado supervisionado e não supervisionado	20
3.3.2 <i>K-Nearest Neighbor</i> - KNN	20
3.3.3 Validação de modelos	22
3.3.4 Trabalhos na literatura com aplicações de aprendizado de máquina na geotecnia.....	23
4. MATERIAIS E MÉTODOS	26
4.1. Banco de dados.....	26
4.2. Metodologia.....	29
4.2.1. Análise descritiva das variáveis	29
4.2.2. Obtenção e validação do modelo.....	30
4.2.3. Comparação do resultado obtido neste trabalho com Santos (2019).....	30
5. RESULTADOS E DISCUSSÕES.....	32
5.1. Análise descritiva do banco de dados.....	32
5.2. Modelo de predição e validação	36
5.3. Comparação com o modelo obtido por Santos (2019)	43
6. CONCLUSÃO.....	44
REFERÊNCIAS BIBLIOGRÁFICAS	45
ANEXO: Banco de dados utilizado no trabalho	50
APENDICE: Script do software R utilizado para a implementação do método KNN utilizando 90% das amostras de teste, semente = 1000 e K = 7.....	54

1. INTRODUÇÃO

A mineração depende de diferentes áreas de conhecimento para que a atividade seja executada de maneira segura e viável. Uma dessas áreas é a geotecnia, que tem por objetivo observar, analisar e prever o comportamento dos solos e rochas, a medida em que são executadas as interferências que o empreendimento mineiro requer sem que ocorram deslizamentos e rupturas.

Dentro da geotecnia, uma área importante é a estabilidade de taludes de cava, que se relaciona aos estudos das condições geotécnicas do solo ou das rochas que compõem o local em suas três escalas, a saber: bancada, inter-rampa e global. Uma ruptura de um talude pode ter consequências severas, uma vez que pode acontecer em uma região que está em operação, podendo levar a perda de vidas, destruição de equipamentos, além de destruição de acessos e comprometimento de todo o trabalho.

Uma maneira de facilitar os trabalhos de geotecnia e minimizar os possíveis erros é o uso de ferramentas de fácil checagem das condições de estabilidade dos taludes da cava em questão, como as características do maciço e o fator de segurança. Segundo Santos (2019), o uso de ferramentas que executam essa função é recomendado, tais como os sistemas propostos por Naghadehi *et al.* (2013) e Santos *et al.* (2019).

Para criação de sistemas de rápida checagem das condições de estabilidade de taludes é fundamental que seja compilado um banco de dados robusto com o máximo possível de variabilidade dos parâmetros, justamente para que vários cenários sejam considerados. O banco de dados, quanto mais completo for e com uma maior quantidade de casos diferentes, gera um aprendizado de máquina mais eficiente e otimizado, com mais informações armazenadas e, conseqüentemente, com uma capacidade maior de prever uma ruptura de um talude, ou executar qualquer que seja sua função.

Um algoritmo simples que é muito utilizado no aprendizado de máquina para fins de proposição de modelos de predição é o KNN (*K-Nearest Neighbor*, ou K Vizinho Mais Próximo), onde o aprendizado acontece com base no estabelecimento das semelhanças que um dado desconhecido tem com outros mais próximos dele, podendo ter várias dimensões. Basicamente o KNN mede a distância entre o dado desconhecido e os dados conhecidos, faz a análise desses dados conhecidos que têm as menores distâncias e a partir daí classifica o dado

desconhecido com base no dado que aparece em maior quantidade entre os mais próximos, ou seja, os vizinhos.

Neste trabalho as amostras do banco de dados foram separadas em amostras de treino e teste para a aplicação do método. Sendo assim, foram obtidos diferentes modelos em que foram medidas as acurácias e, por fim, o resultado foi comparado com o obtido por Santos (2019).

2. OBJETIVOS

2.1 Objetivo geral

Aplicar o método *K-Nearest Neighbors* para proposição de um modelo de predição da condição de estabilidade de taludes rochosos de mina.

2.2 Objetivos específicos

- Obter o modelo de predição da condição de estabilidade de taludes de mina;
- Comparar o modelo e os resultados obtidos com o modelo obtido por Santos (2019).

3. REVISÃO BIBLIOGRÁFICA

3.1. Conceitos iniciais em estabilidade de taludes de cava

A inclinação do talude é definida a partir das características geomecânicas do material no qual a obra será executada, de forma que o avanço operacional da mina possa ocorrer de forma segura e econômica. A estabilidade de taludes de cava está relacionada a estratégias de maximização de lucro, além de possíveis situações de risco levando em consideração diversos aspectos que envolvem o local. Os principais pontos de análise em projetos de taludes de mina a céu aberto são o ângulo de inclinação global (podendo variar ao redor da mina para se ajustar a diferentes características geológicas), da crista à base, o ângulo do talude inter-rampas e o ângulo de face das bancadas. Além disso, os principais fatores que podem influenciar a construção do talude são a sua altura, geologia, resistência da rocha e pressões de água subterrânea (Wyllie & Mah, 2004).

Os principais parâmetros a serem considerados são as características geomecânicas, os parâmetros geométricos, além das variáveis externas, tanto naturais, como alterações climáticas, quanto artificiais, como as detonações dos desmontes a fogo (Santos *et al.* 2018). Podem ocorrer três escalas de rupturas nos taludes em minas a céu aberto: *bench slope failure* (ruptura de bancada), *inter-ramp slope failure* (ruptura entre rampas) e *overall slope failure* (ruptura global). Essas rupturas estão relacionadas ao volume do material rompido, à escala da ruptura e, conseqüentemente, ao potencial de dano que pode ser causado. Segundo Sjoberg (1996), em uma escala de bancada, por exemplo, algumas rupturas controladas, como rupturas em cunha, são comuns. Porém, quando isso acontece em uma escala maior, essas rupturas simples (com controle estrutural claro) passam a ser menos recorrentes e rupturas mais complexas podem surgir. Segundo Santos (2019), cada tipo de escala de ruptura de taludes pode se caracterizar como:

- Rupturas de bancada: apresenta um impacto mínimo, tendo, em sua maior parte, custos relacionados à limpeza do material. Esse tipo de ruptura é relativamente comum e é inevitável que aconteça durante a operação, principalmente por eventos naturais ou até mesmo pela execução da atividade;
- Rupturas em taludes inter-rampa: nesse tipo de ruptura os transtornos são bem mais consideráveis e graves, como, por exemplo, danos de equipamentos e possíveis ferimentos de operários, devido ao volume maior de material. Por consequência do aumento dos danos, o

prejuízo econômico também será muito maior, não só pela limpeza do material, como também para a reparação dos demais problemas que são gerados;

- Rupturas de talude global: é o tipo de escala que leva a consequências mais graves e exige um cuidado maior para que os danos sejam mitigados caso aconteça, uma vez que tem uma grande chance de danificar muitos equipamentos e causar lesões sérias e até a morte dos operários. Além disso, o impacto econômico pode ser irreversível, uma vez que o minério pode ser diluído, o que leva a perda do valor econômico.

A Figura 1 ilustra os tipos de escala dos taludes de mina. Nota-se que, levando em consideração as diferentes escalas, existe diferença no volume de material rompido e, por isso, as consequências variam tanto de uma escala de ruptura para outra.

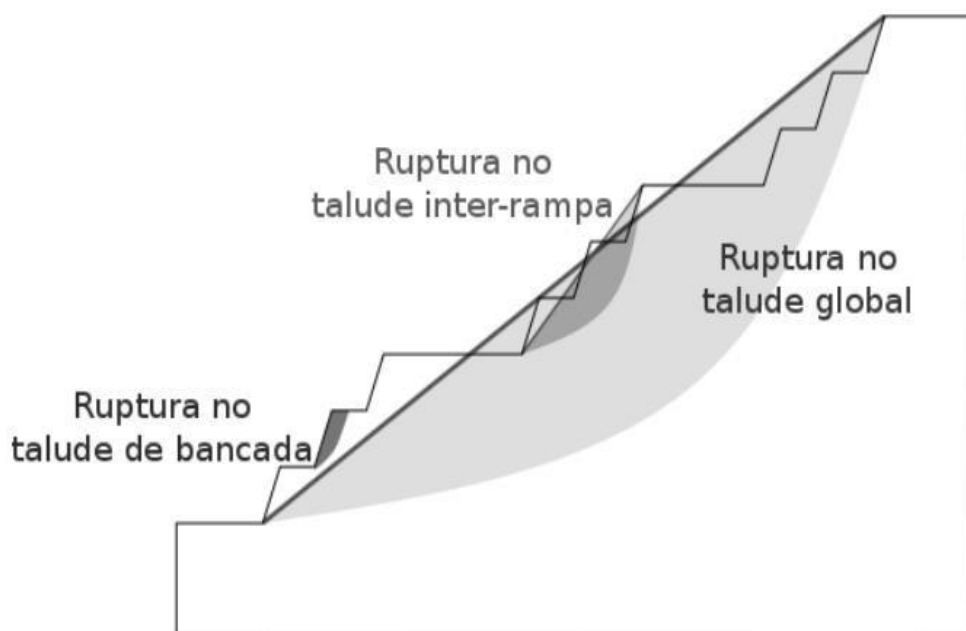


Figura 1 - Escalas de taludes de mina (Santos, 2019).

3.2. Maciços rochosos

Rocha pode ser definida como uma associação natural de minerais em proporções definidas que ocorre em uma extensão considerável, podendo ser agrupadas em três grandes grupos: ígneas, metamórficas ou sedimentares (Branco, 2015). Para Mesquita (2008), rocha pode ser definida como um corpo sólido e natural formado por um conjunto de grãos de minerais resultados a partir de processos geológicos que ocorrem sob pressão e temperatura específicas.

O conceito de maciço rochoso, para Fiori & Carmignani (2013), pode ser dado como uma massa de rocha constituída por blocos discretos com propriedades de rochas intactas, interrompida por descontinuidades.

As descontinuidades dos maciços rochosos, segundo Santos (2021), formam os planos de fraqueza das rochas e podem ser de diferentes tipos, desde juntas, fraturas, planos de acamamento e planos de foliação. Essas descontinuidades são fundamentais de serem estudadas para compreender o maciço rochoso, uma vez que influenciam significativamente em sua resistência e seu comportamento geomecânico.

As descontinuidades podem ser encontradas tanto de maneira difusa, próximas ou afastadas. Além disso, elas podem apresentar diferentes tipos e quantidades de materiais de preenchimento. Quando essas descontinuidades se encontram no maciço com a mesma orientação, elas pertencem à mesma família e um conjunto dessas famílias é definido como um sistema (Jaques, 2014).

Na Figura 2 é representado um bloco diagrama que exemplifica as principais características das descontinuidades para avaliar o maciço rochoso.

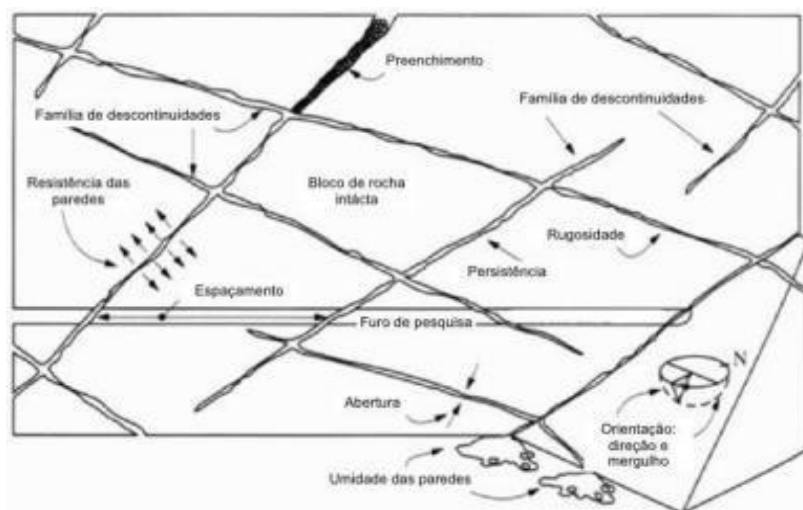


Figura 2 - Apresentação dos principais parâmetros geomecânicos avaliados na caracterização do maciço rochoso (LACERDA, S.G. et al., 2019).

Alguns conceitos são fundamentais no estudo dos maciços rochosos, uma vez que eles são os parâmetros que definem as características desses maciços e, conseqüentemente, o seu comportamento. Alguns dos conceitos a serem citados são os de espaçamento, persistência, rugosidade, preenchimento, abertura e orientação. Segundo Filho & Nummer (2014), esses termos podem ser definidos como:

- **Espaçamento:** pode ser definido como a distância entre as descontinuidades adjacentes de uma mesma família, podendo ser medido em linha de visada na face do talude. A Figura 3 ilustra uma medição aparente do espaçamento de descontinuidades. Para a medida real, o espaçamento deve ser tomado na linha ortogonal às descontinuidades adjacentes da mesma família.

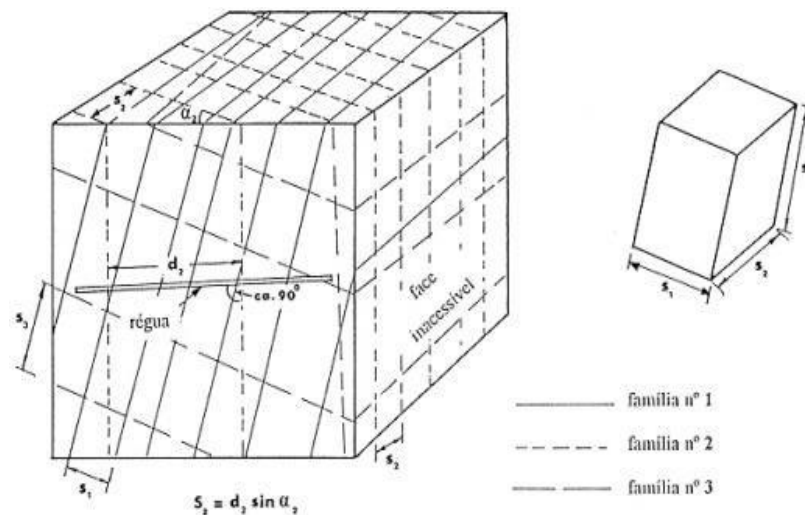


Figura 3 - Medição de espaçamento de descontinuidades em uma face exposta (Filho & Nummer, 2014).

- **Persistência:** A persistência no comprimento do traço de uma descontinuidade e tem grande influência no maciço rochoso. A Tabela 1 mostra a classificação das persistências ou descontinuidades.

Continuidade	Comprimento (m)
Muito pequena	< 1
Pequena	1 - 3
Média	3 - 10
Elevada	10 - 20
Muito elevada	> 20

Tabela 1 - Classificação das persistências (Filho e Nummer, 2014).

- **Rugosidade:** Como o próprio nome sugere, a rugosidade quantifica as ondulações presentes nas paredes das descontinuidades abertas e vai influenciar na resistência

ao cisalhamento ao longo da descontinuidade. A caracterização da rugosidade é dada pelas irregularidades nas paredes da descontinuidade pela sua curvatura e tem um parâmetro que a quantifica, o JRC, que foi proposto por Barton e Choubey e é representado na figura 4.

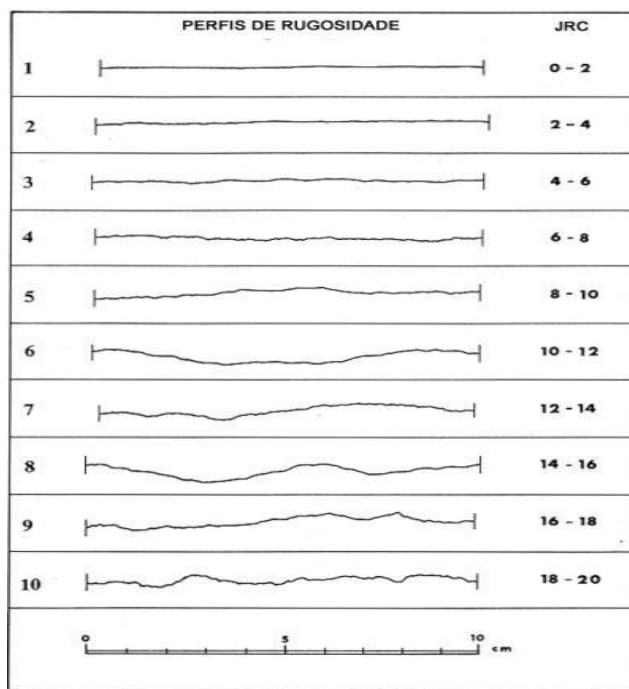


Figura 4 - Perfis de rugosidade e seus valores de JRC (Adaptado de Barton & Choubey, 1977).

- **Preenchimento:** pode ser definido com o material que preenche as descontinuidades da rocha, podendo variar quanto às suas características e, assim, influenciar no comportamento do maciço rochoso. O preenchimento pode ser definido como o material que fica entre as paredes da descontinuidade, sendo comumente mais fraco que a rocha original e costuma ser arenoso ou argiloso (Wolpp, 2018). Esse preenchimento pode ser mais ou menos consolidado, uma vez que essa consolidação pode ser fundamental à resistência ao longo da descontinuidade.
- **Abertura:** é a distância perpendicular entre duas paredes paralelas de uma descontinuidade e pode estar preenchida, normalmente, por água ou ar (Wolpp, 2018). A tabela 2 determina a classificação das aberturas.

Descrição	Abertura
Muito fechadas	< 0,1 mm

Fechadas	0,1 - 0,25 mm
Parcialmente abertas	0,25 - 0,5 mm
Abertas	0,5 - 2,5 mm
Cavernosas	> 1 m

Tabela 2 - Classificação das aberturas (Adaptado de ISRM, 1978).

- **Orientação:** pode ser definida como a orientação da descontinuidade no espaço, levando em consideração o mergulho e a direção de mergulho do plano de fraqueza (Wolpp, 2018). De acordo com Filho & Nummer (2014), a orientação é o mergulho da linha de maior declive do plano que ela está contida, a partir do ângulo medido no sentido descendente a partir da horizontal e pelo azimuth da direção dessa linha. O ângulo é medido a partir do Norte no sentido horário, como ilustrado na figura 5.

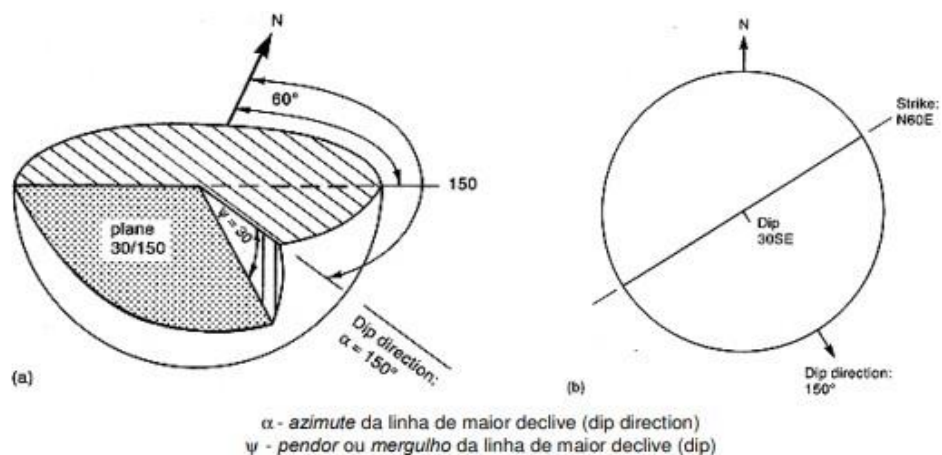


Figura 5 - Definição da orientação do plano de uma descontinuidade N60E,30SE: (a) vista simétrica; (b) vista em planta (Filho & Nummer, 2014).

As características do maciço rochoso impactam diretamente em sua qualidade e, consequentemente, em suas propriedades de resistência mecânica. Isso pode ser verificado através das classificações geomecânicas que são utilizadas para avaliar a qualidade dos maciços rochosos. Entre elas, pode-se citar: RMR (*Rock Mass Rating*), RQD (*Rock Quality Designation*), Sistema - Q (*Rock Tunneling Quality*) e o GSI (*Geological Strength Index*). Esses métodos podem ser definidos como:

- RQD: Quantifica a qualidade das rochas dos testemunhos de sondagem baseando-se na percentagem de recuperação, considerando apenas partes do testemunho com mais de 10 cm de comprimento.

Além disso, Deere (1989) estabelece uma classificação de maciços rochosos utilizando o RQD. Essa classificação é ilustrada na tabela 3.

Classe	Qualidade da rocha	RQD (%)
I	Muito ruim	<25
II	Ruim	25 - 50
III	Regular	50 - 75
IV	Bom	75 - 90
V	Excelente	90 - 100

Tabela 3 - Classes da qualidade de classificação do maciço rochoso de acordo com o RQD (Adaptado de Deere, 1989).

- RMR: Foi criado por Bieniawski (1989) e faz uso de alguns parâmetros para a avaliação do maciço rochoso, são alguns dos mais importantes dos parâmetros: a resistência à compressão uniaxial da rocha, o espaçamento entre as descontinuidades, a condição das descontinuidades, a condição de percolação d'água, a orientação das descontinuidades, além do próprio RQD citado anteriormente.

Além disso, Bieniawski criou uma tabela (como está ilustrado na tabela 4) onde esses maciços rochosos podem ser classificados segundo cinco classes, sendo os cinco primeiros listados a seguir, já o sexto parâmetro está ligado ao ajuste da orientação das descontinuidades. Cada uma das classes tem um conjunto de 20 valores, então o RMR pode variar entre 0 e 100 (adimensional).

Classificação do maciço rochoso					
Somatório	100 - 81	80 - 61	60 - 41	40 - 21	< 21
Classe	I	II	III	IV	V

Descrição da rocha	Muito boa	Boa	Razoável	Pobre	Muito pobre
--------------------	-----------	-----	----------	-------	-------------

Tabela 4 - Classificação de maciços rochosos segundo o RMR (Adaptado de Bieniawski, 1989).

- Q: Faz uso de parâmetros para a classificação dos maciços rochosos, variando numa escala logarítmica de 0,001 a 1000, quantificando a qualidade destes. Os 6 parâmetros são: o RQD novamente, o número de famílias de descontinuidades (J_n), o índice de rugosidade das descontinuidades (J_r), o grau de alteração das descontinuidades (J_a), o fator de redução da água (J_w) e o estado de tensão do maciço (SFR).

A partir desse valor calculado do sistema Q, é utilizada a tabela 5 para classificar o maciço em 9 classes:

Classe	Padrão geomecânico do maciço	Valores de Q
IX	Péssimo	0,0001 - 0,01
VIII	Extremamente ruim	0,01 - 0,1
VII	Muito ruim	0,1 - 1,0
VI	Ruim	1,0 - 4,0
V	Regular	4,0 - 10,0
IV	Bom	10,0 - 40,0
III	Muito bom	40,0 - 100,0
II	Ótimo	100,0 - 400,0
I	Excelente	> 400,0

Tabela 5 - Classificação do maciço rochoso pelo sistema Q (Adaptado de Bieniawski, 1989).

- GSI: Foi desenvolvido inicialmente para maciços homogeneamente fraturados e possibilita a estimativa de propriedades mecânicas dos maciços e a redução da força do maciço para diferentes condições geológicas. A figura 6 indica os parâmetros para determinação do valor do GSI.

ÍNDICE DE RESISTÊNCIA GEOLÓGICA PARA MACIÇOS FRATURADOS (Hoek & Marinos, 2000) A partir da litologia, estrutura e condições da superfície das descontinuidades, estima-se o valor médio de GSI. Não tente ser tão preciso. Citar um alcance de 33 a 37 é mais realístico do que afirmar que GSI = 35. Observe que a tabela não é aplicável a rupturas controladas estruturalmente. Onde existir planos estruturais de fraqueza com orientação desfavorável em relação à face da escavação, estes controlarão o comportamento do maciço rochoso. A resistência ao cisalhamento de superfícies rochosas propensas a sofrer deterioração com as variações de umidade será reduzida quando existir presença de água. Quando trabalhar com rochas de categoria razoável a muito pobre, um deslocamento para a direita pode ser feito em condições de saturação. Pressão de água é avaliada através de análises de tensão efetiva.		CONDIÇÕES DA SUPERFÍCIE DECRESÇIMO DA QUALIDADE DA SUPERFÍCIE →				
ESTRUTURA		MUITO BOA Muito rugosa, superfícies não interperisadas	BOA Rugosa, levemente interperisada, superfícies manchadas de ferro ("enferrujadas")	RAZOÁVEL Lisa, superfícies moderadamente interperisadas e alteradas	POBRE Polidas, superfícies altamente interperisadas com cárnadas compactas ou preenchimentos ou fragmentos angulares	MUITO POBRE Polidas, superfícies altamente interperisadas com cárnadas de argila mole ou preenchimentos
 INTACTA OU MACIÇA – rocha intacta ou maciça com poucas descontinuidades, largamente espaçadas		90	80	70	N/A	N/A
 FRATURADA – maciço rochoso bem intertravado não perturbado constituído por blocos cúbicos formados por três famílias de descontinuidades que se interceptam			90	80		
 MUITO FRATURADA – intertravado, maciço parcialmente perturbado com blocos angulares com várias faces formados por quatro ou mais famílias de descontinuidades				90		
 FRATURADA/PERTURBADA "COSTURADA" – dobrada com blocos angulares formados por muitas famílias de descontinuidades interceptadas. Persistência de planos de acamamentos ou xistoides				40	30	
 DESINTEGRADA – pobremente intertravada, maciço rochoso intensamente quebrado, com uma mistura de pedaços rochosos angulares e arredondados					20	
 LAMINADA/CISALHADA – ausência de blocos devido ao estreito espaçamento de xistoides fracas ou planos de cisalhamento		N/A	N/A			10

Figura 6 – parâmetros para determinação do GSI (Adaptado de Hoek & Marinos, 2000)

3.3. Aprendizado de máquina

O aprendizado de máquina, de acordo com Goodfellow *et al.* (2016), é uma forma de estatística aplicada em computadores com o objetivo de estimar funções complexas e permite que sejam desenvolvidas tarefas difíceis de resolver com programação. A partir de informações inseridas pelo usuário no banco de dados, a máquina consegue identificar padrões para diferentes situações e, assim, projetar soluções a partir das características em comum com as que foram inseridas inicialmente. Sendo assim, nota-se que é fundamental que se tenha um banco de dados robusto e completo, uma vez que isso torna o aprendizado de máquina mais eficiente.

O processo de aprendizado de máquina pode ser dividido em diferentes fases, desde o treinamento até que sejam obtidas as predições. Nesta fase inicial, a máquina recebe as informações de como selecionar, classificar e coletar dados é chamada de fase de treinamento, onde os processos são formulados para executar o trabalho. Na figura 7 é esquematizada a

sequência do aprendizado de máquina, passando pela fase de treinamento, o classificador e as previsões obtidas.

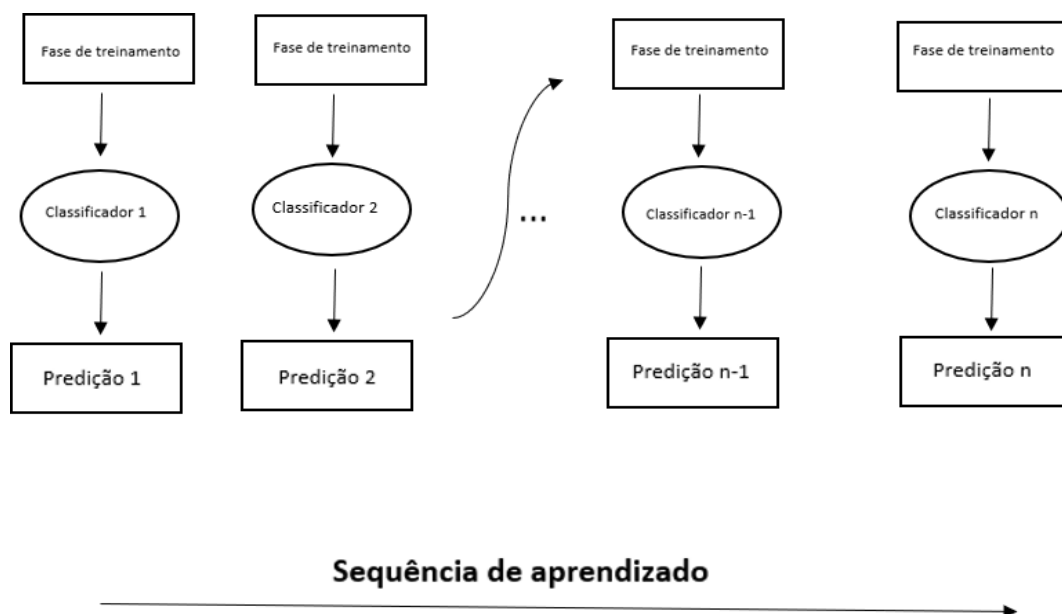


Figura 7 - Fluxograma de sequência de aprendizagem de máquina (Adaptado de Pham, et al. 2021).

3.3.1 Aprendizado supervisionado e não supervisionado

Segundo Goodfellow *et al.* (2016), o aprendizado de máquina supervisionado pode ser definido como algoritmos que associam as entradas com alguma saída a partir de um conjunto de exemplos de entradas e saídas, entretanto em alguns casos as saídas podem ser difíceis de serem obtidas automaticamente e se faz necessário seu fornecimento pelo usuário.

O aprendizado de máquina não supervisionado não vai ter uma saída fornecida, uma vez que estão sendo buscados “grupos” e apesar de ser pouco utilizada, é muito eficiente para guiar estudos na exploração de dados para análises. Ainda segundo Goodfellow *et al.* (2016), o aprendizado de máquina não supervisionado utiliza apenas recursos para obter seus dados, não sendo necessária a interferência do usuário e é comumente utilizado para encontrar a melhor representação de dados levando em consideração todas as restrições pré-definidas.

3.3.2 *K-Nearest Neighbor* - KNN

O KNN ou *K-Nearest Neighbors*, segundo Mariz (2017), pode ser definido como um método de aprendizado supervisionado de predição que utilizará a distância entre a amostra atual e o número de vizinhos próximos previamente definido pela variável k e, assim, definir

qual o resultado da sua predição. O funcionamento do algoritmo ocorre da seguinte forma: a partir do parâmetro k , levando-se em consideração um grupo de amostras, o usuário consegue descobrir a classe de uma das amostras comparando a sua distância com todas as outras amostras do grupo, sendo que essa distância pode ser calculada de diferentes formas. Esse método, devido ao seu custo operacional, é mais utilizado (e mais viável) em situações que existem poucos parâmetros.

A distância euclidiana é a distância mais utilizada para a finalidade de calcular a distância entre dois pontos. De acordo com Silva (2005), a distância que comumente é utilizada no cálculo do método KNN é a euclidiana, que é calculada de acordo com a Equação (1), onde $X = (x_1, x_2, \dots, x_n)$ e $Y = (y_1, y_2, \dots, y_n)$ são dois pontos do $\mathbb{R}^n$.

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

Onde:

$d(x, y)$ é a distância entre os pontos x e y ;

x_i é a coordenada i do ponto x ;

y_i é a coordenada i do ponto y ;

De acordo com Pham *et al.* (2021), o KNN é um algoritmo de *Machine Learning* não paramétrico que se caracteriza por ter uma “fase de treinamento” alta e um alto custo operacional para a máquina, o que significa que, devido à grande entrada de dados, o processamento será relativamente complexo. Além disso, a velocidade e o desempenho do método estarão diretamente relacionados com a quantidade de dados de entrada e a quantidade de “vizinhos próximos”.

A figura 8 ilustra o método KNN em gráfico 2D. Nota-se que existem 2 classes diferentes classificadas nos eixos x e y , e existe uma incógnita no meio. 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 analisados levando em consideração os parâmetros definidos.

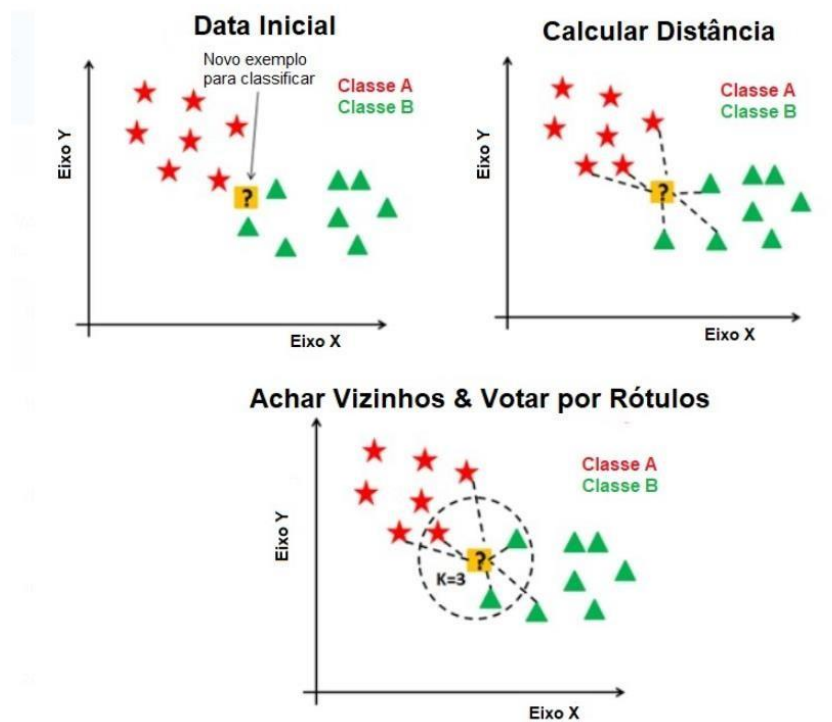


Figura 8 - Modelo gráfico 2D do funcionamento do método KNN (Adaptado de Azank, 2019).

3.3.3 Validação de modelos

A validação cruzada, de acordo com Oliveira (2016), tem a função de dividir o conjunto de dados inseridos em uma determinada quantidade de estratos independentes de dados (*folds*) e, assim, utilizar um desses estratos como conjunto de teste enquanto os outros são usados para treinamento. A execução acontece variando o estrato que é utilizado para teste para que no fim do processo todos os estratos sejam utilizados para teste uma vez. Esse processo é nomeado com *k-fold cross-validation* onde k é o número de estratos independentes e, ao fim do processo,

são resultados, k modelos construídos com conjuntos distintos de dados e testados com conjuntos de dados diferentes de treinamento. Os objetivos de realizar esse processo são: escolher a configuração mais adequada aos dados que se tem no esquema de aprendizado de máquina e estimar o desempenho que um modelo teria se fosse aplicado com dados diferentes.

Existem diferentes formas de avaliar uma classificação, uma delas é utilizando a matriz de confusão, que, de acordo com Passos (2013), pode ser definida como uma matriz quadrada que relaciona o número de unidades de uma amostra que pertence a uma categoria particular de duas classificações. As linhas dessa matriz são definidas de acordo com as classificações do método utilizado enquanto as colunas representam os dados de referência, que, no fim, vão

definir a verdade de campo. Ainda de acordo com Passos (2013), a frequência na diagonal (n_{11} , n_{22} , ..., n_{ii}) vai indicar os elementos que estão classificados da forma certa.

É ilustrado na tabela 6 um exemplo de matriz de confusão indicado com os valores de n_{11} , n_{12} , ..., n_{33} onde as classes são “ST”, “FSB” e “OF” (Santos, 2019).

Classe real	Classificação obtida a partir da regra de classificação				
	População	ST	FSB	OF	Total
	ST	n_{11}	n_{12}	n_{13}	n_1
	FSB	n_{21}	n_{22}	n_{23}	n_2
	OF	n_{31}	n_{32}	n_{33}	n_3

Tabela 6 - Exemplo de matriz de confusão para três classes (Santos, 2019)

Para fazer o cálculo da probabilidade global de acerto (PGA) e da taxa de erro aparente (TEA) são utilizadas as equações (2) e (3) para uma matriz de confusão de três classes.

$$PGA = \frac{n_{11} + n_{22} + n_{33}}{n_1 + n_2 + n_3} \quad (2)$$

$$TEA = \frac{n_{12} + n_{21} + n_{31} + n_{13} + n_{23} + n_{32}}{n_1 + n_2 + n_3} \quad (3)$$

3.3.4 Trabalhos na literatura com aplicações de aprendizado de máquina na geotecnia

O artigo “*An Improved KNN-Based Slope Stability Prediction Model*” (Huang *et al*, 2020), que, traduzindo o seu título para o português, significa “Um modelo de previsão de estabilidade de taludes baseado em KNN”, faz uso do WKNN (Onde o W é sigla da palavra “*weighted*” que significa pesado, o que implica que no cálculo da distância dos pontos, eles receberão pesos diferentes) para definir a inclinação dos taludes, de forma que, a partir do banco de dados inserido, o método vai criar soluções a partir das características em comum dos dados obtidos. Para a realização desse estudo, foram considerados os parâmetros dos taludes: ângulo do talude, altura do talude, coesão, ângulo de atrito interno, peso específico, nível de água subterrânea, intensidade dos tremores de terra e a intensidade da chuva. Sendo assim, os autores utilizaram um banco de dados de 50 amostras de treinamento e 14 amostras de teste de estabilidade de taludes e obtiveram uma precisão de acurácia de 92,30%.

Outro artigo que pode ser mencionado sobre a utilização do aprendizado de máquina na estabilidade de taludes é “*Predicting Slope Stability Failure through Machine Learning*

Paradigms” de Bui *et al.* (2019), em que os autores tratam da previsão de falha da estabilidade de taludes utilizando o aprendizado de máquina utilizando diferentes técnicas para prever o fator de segurança dos taludes. Alguns desses métodos utilizados são: Rede Neural *Perceptron* Multicamadas “*multi-layer perceptron*” (MLP), Regressão do Progresso “*Gaussian process regression*” (GPR), Regressão Linear Múltipla “*multiple linear regression*” (MLR), Regressão Linear Simples “*simple linear regression*” (SLR) e Regressão de Vetor de Suporte “*support vector regression*” (SVR). O estudo tem o objetivo de avaliar e otimizar modelos de regressão multilinear e utilizar o aprendizado de máquina para prever os fatores de segurança. Para esse estudo foi utilizado um banco de dados de 630 taludes, sendo 504 de treinamento e 126 de teste e os parâmetros utilizados para análise são: resistência ao cisalhamento não drenado (C_u), ângulo de talude (β), razão de distância de recuo (b/B), e sobretaxa aplicada na sapata instalada sobre o talude (w). A implementação dos modelos foi feita no software WEKA e, fazendo os cálculos utilizando o R^2 (coeficiente de determinação), MAE (erro absoluto médio), RMSE (erro quadrático médio), RAE (erro absoluto relativo) e RRSE (erro quadrático relativo), foram obtidos os resultados: na fase de treinamento, R^2 (0,9467, 0,9586, 0,9937, 0,99019 e 0,9529), MAE (1,5598, 1,2527, 0,4940, 1,7013 e 1,161), RMSE (1,9957, 1,7366, 0,7131, 2,6334 e 1,9183), RAE (31,1929%, 25,0515%, 9,8796%, 34,0224%, e 23,2182%) e RRSE (32,7404%, 28,4887%, 11,6985%, 43,2016% e 31,4703%), respectivamente para GPR, MLR, MLP, SLR e SVR e, para a fase de teste, R^2 (0,9509, 0,9649, 0,9939, 0,9265 e 0,9653), MAE (1,5291, 1,1949, 0,5155, 1,5387 e 1,0364), RMSE (1,9447, 1,5891, 0,7039, 2,2618 e 1,6362), RAE (30,4047, 24,4047 %, 10,4047%, 31,0892% e 20,9366%) e RRSE (32,3841%, 26,4613%, 11,8116%, 37,6639% e 27,2470%), respectivamente para GPR, MLR, MLP, SLR e SVR, o que implicou que o MLP tem uma maior eficiência.

Moayedi *et al.* (2019) também escreveram sobre a estabilidade de taludes utilizando aprendizado de máquina no artigo “*Machine-Learning-Based Classification Approaches towards Recognizing Slope Stability Failure*” onde investigam a aplicabilidade da combinação de modelos de aprendizado de máquina na avaliação da estabilidade de taludes usando, também, a Regressão Linear Múltipla (MLR), a Rede Neural *Perceptron* multicamadas (MLP), além de Regressão da Função de Base Radial “*radial basis function regression*” (RBFR), Máquina de Vetor de Suporte Aprimorada Usando Algoritmo de Otimização Mínima Sequencial “*improved support vector machine using sequential minimal optimization algorithm*” (SMO-SVM), K-Vizinho Próximo Lento “*lazy k-nearest neighbor*” (IBK), Floresta Aleatória “*random forest*”

(RF), e Árvore Aleatória “*random tree*” (RT). Esses métodos são usados com o objetivo de avaliar a estabilidade de taludes a partir da estimativa do fator de segurança e, nesse artigo, foi feita uma classificação comparativa com base em cinco categorias de estabilidade de taludes. Os parâmetros utilizados também foram resistência ao cisalhamento não drenado, ângulo de talude, razão de distância de recuo, e sobretaxa aplicada na sapata instalada sobre o talude e, também, foram utilizados 630 taludes no banco de dados, sendo 504 de treinamento e 126 de teste. Os resultados obtidos foram: R^2 (0,9586, 0,9937, 0,9948, 0,9529, 1,000, 0,9997 e 0,9998), RMSE (28,4887, 1,7366, 0,7131, 0,6231, 1,9183, 0,0000, 0,1486 e 0,1111), e RRSE (11,6985%, 10,2221%, 31,4703%, 0,0000%, 2,4385% e 1,8221%) para os dados de treinamento e R^2 (0,9649, 0,9939, 0,9955, 0,9653, 0,9837, 0,9985, e 0,9950), MAE (1,1939, 0,5149, 0,3936, 1,0360, 0,8184, 0,2152 e 0,3492) e RAE (24,1272%, 10,4047%, 7,9549%, 20,9366%, 16,5388%, 4,3498% e 7,0576%) para os dados de teste, obtendo-se, dessa forma, a superioridade do RF.

Santos (2019) fala em “Análise de risco geotécnico em taludes rochosos de mina com uso de técnicas estatísticas multivariadas e de aprendizado de máquina” sobre métodos para analisar riscos na estabilidade de taludes rochosos utilizando técnicas de estatística multivariada e aprendizado de máquina na qual ela propôs um modelo de predição para taludes rochosos de mina para a previsão de condição de estabilidade. Para a realização do trabalho, foi utilizado um banco de dados constituído de 88 taludes de mina, as variáveis utilizadas foram a resistência à compressão uniaxial da rocha intacta, o espaçamento, a persistência, a abertura, a rugosidade, o preenchimento, a orientação da descontinuidade principal, a alteração da rocha, a condição de água subterrânea, o método de desmonte, a altura e a inclinação do talude global, as classes da variável classificatória foram: talude estável, talude com ruptura inter-rampa, talude com ruptura global. Além disso, a autora aplicou análise discriminante para o modelo capaz de prever a condição de estabilidade de taludes e a função apresentou erro de 11,36%.

4. MATERIAIS E MÉTODOS

4.1. Banco de dados

O banco de dados utilizado neste trabalho foi desenvolvido por Nagadehi et al. (2013) e por Santos (2019) e é composto por 88 taludes de mina ao redor do mundo, que são apresentados na tabela 7.

Mina	País	Minério	Tipo de rocha (litologia)	Número de taludes em cada mina
Águas Claras	Brasil	Ferro	Dolomito, xisto, filito e quartzito	5
Aitik	Suécia	Cobre	Gnaiss, diorito, xisto	6
Alegria	Brasil	Ferro	Itabirito, xisto e dolomito	4
Angooran	Irã	Chumbo e zinco	Calcário	4
Aznalcollar	Espanha	Chumbo e zinco	Xisto e filito	5
Betze-Post	EUA	Ouro	Calcário e diorito	4
Cadia Hill	Austrália	Ouro e cobre	Diorito	5
Chadormalou	Irã	Ferro	Diorito	5
Choghart	Irã	Ferro	Filito e xisto	5
Chuquicamata	Chile	Cobre	Granodiorito e granito	5
Escondida	Chile	Cobre	Andesito e diorito	7
Esperanza	EUA	Cobre	Andesito	1
Gole-Gohar	Irã	Ferro	Xisto, gnaiss e hematita	4
La Yesa	Espanha	Argilito	Arenito e conglomerado	2
Ok Tedi	Papua Nova Guiné	Ouro e cobre	Silito	2
Panda	Canadá	Diamante	Granodiorito	1
Sandsloot	África do Sul	Platina	Norito, piroxenito e gabro	6

Sarcheshmeh	Irã	Cobre	Andesito	4
Sungun	Irã	Cobre	Monzonito e diorito	5
Ujina	Chile	Cobre	Riolito e andesito	1
Venetia	África do Sul	Diamante	Xisto, gnaisse e quartzito	7

Tabela 7 - Minas do banco de dados (Santos, 2019).

As variáveis utilizadas para a construção do banco de dados são:

- P1: Resistência à compressão uniaxial;
- P2: Espaçamento;
- P3: Persistência da descontinuidade principal;
- P4: Abertura da descontinuidade principal;
- P5: Rugosidade da descontinuidade principal;
- P6: Preenchimento da descontinuidade principal;
- P7: Alteração da rocha;
- P8: Condição de percolação de água subterrânea;
- P9: Orientação relativa da descontinuidade principal;
- P10: Método de desmonte empregado na mina;
- P11: Altura do talude global;
- P12: Inclinação do talude global.

As variáveis podem ser divididas entre as que têm caráter quantitativo e as que têm caráter qualitativo, que foram avaliadas com pesos de acordo com as melhores condições geotécnicas. Além disso, cada talude do banco de dados é avaliado, por fim, como estável (ST) ou instável (ruptura inter-rampa FSB e ruptura global OF), os valores associados aos parâmetros são apresentados na tabela 8.

Parâmetro/ Variável	Tipo da variável
UCS - Resistência à compressão uniaxial da rocha intacta (MPa) – P1	Quantitativa
Espaçamento (m) – P2	Quantitativa
Persistência da descontinuidade principal (m) – P3	Quantitativa

Abertura da descontinuidade principal (mm) – P4						Quantitativa
Rugosidade da descontinuidade principal (Adaptado de Bieniawski, 1989) – P5						Qualitativa ordinal
Parâmetro	Muito rugosa	Rugosa	Levemente rugosa	Lisa	Polida	
Peso	5	4	3	2	1	
Preenchimento da descontinuidade principal (Adaptado de Bieniawski, 1989) – P6						Qualitativa ordinal
Parâmetro	Nenhum	Preenchimento duro		Preenchimento macio		
		< 5 mm	> 5 mm	< 5 mm	> 5 mm	
Peso	5	4	3	2	1	
Alteração da rocha (ISRM, 2014) – P7						Qualitativa ordinal
Parâmetro	Não alterada (W1)	Ligeiramente alterada (W2)	Moderadamente alterada (W3)	Muito alterada (W4)	Decompоста (W5)	
Peso	5	4	3	2	1	
Condição de água subterrânea (Adaptado de Bieniawski, 1989) – P8						Qualitativa ordinal
Parâmetro	Seco	Úmido	Encharcado	Gotejando	Com fluxo	
Peso	5	4	3	2	1	
Orientação da descontinuidade principal (Adaptado de Naghadehi, 2013) – P9						Qualitativa ordinal
Parâmetro	$\beta d > \beta s$ $\alpha d - \alpha s > 30^\circ$	$\beta d > \beta s$ $\alpha d - \alpha s < 30^\circ$	$0 \leq \beta d \leq \beta s/4$ $\alpha d - \alpha s > 30^\circ$	$\beta s/4 \leq \beta d \leq \beta s/2$ $\alpha d - \alpha s < 30^\circ$	$\beta s/2 \leq \beta d \leq \beta s$ $\alpha d - \alpha s < 30^\circ$	
Descrição	Muito favorável	Favorável	Razoável	Desfavorável	Muito desfavorável	
Peso	5	4	3	2	1	
Método de desmonte (Adaptado de Naghadehi, 2013) – P10						Qualitativa ordinal
Parâmetro	Pré-fissuramento	Pós-fissuramento	Smooth wall / cushion	Modified production blast	Desmonte regular / mecânico	
Peso	5	4	3	2	1	
Altura do talude global (m) – P11						Quantitativa
Inclinação do talude global (°) – P12						Quantitativa

Tabela 8 - Valores associados aos parâmetros (Santos, 2019).

Considerar o “ α_d ” como azimuth do mergulho da descontinuidade principal, α_s como azimuth do mergulho do talude, β_d como mergulho da descontinuidade principal e β_s como mergulho do talude.

A tabela 9 apresenta os dez primeiros taludes do banco de dados com seus parâmetros, a condição de estabilidade e a escala de ruptura (sendo “ST” estável, “FSB” ruptura no talude inter-rampa e bancada e “OF” ruptura no talude global).

Talude	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	condição de estabilidade
1	101	0,22	2,25	1,25	3	3	4	5	3	1	75	5	ST
2	70	1,50	6,5	3	4	3	4	5	2	2	20	5	ST
3	60	2,50	17,5	3	3	1	4	4	3	2	20	5	FSB
4	60	3,00	17,5	3	3	2	4	4	4	2	00	5	FSB
5	47	1,30	15	3	3	2	3	5	2	1	60	5	ST
6	62	2,00	6,5	6	4	2	4	4	3	1	80	5	ST
7	42	1,20	15	3	2	2	2	4	3	1	490	5	FSB
8	57	1,20	12,5	3	3	3	3	4	3	1	340	5	ST
9	22	0,82	0,91	3	3	3	3	3	5	5	96	6	OF
10	114	0,31	0,5	3	3	4	3	4	2	5	180	55	ST

Tabela 9 - Taludes do banco de dados (Santos, 2019).

4.2. Metodologia

4.2.1. Análise descritiva das variáveis

A análise descritiva das variáveis, de acordo com Reis & Reis (2002), 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, quando podem ser separadas por categorias em vez de valores quantitativos.

Para as variáveis quantitativas, as métricas de média simples, valor máximo, valor mínimo e desvio padrão foram obtidas e estão apresentadas no capítulo de resultados e discussão. No caso das variáveis qualitativas, as frequências das categorias de cada variável foram obtidas e apresentadas em forma de gráfico de pizza.

4.2.2. Obtenção e validação do modelo

O KNN foi utilizado para gerar o modelo preditor, onde, inicialmente, o banco foi dividido em amostras de treino e teste de maneira aleatória. A amostra de treino foi utilizada para gerar o modelo e a amostra de teste foi utilizada para a sua validação, sendo possível readequar a proporção da quantidade de amostras de treino e de teste. Para isso, as métricas de PGA (probabilidade global de acerto) e de TEA (taxa de erro aparente) foram utilizadas e, para a obtenção e validação do modelo, foi utilizado o software livre R (RCORE Team, 2022).

4.2.3. Comparação do resultado obtido neste trabalho com Santos (2019)

O modelo e as métricas de validação obtidas neste trabalho foram comparadas com os resultados obtidos por Santos (2019) com o objetivo de determinar se o método KNN forneceu um modelo mais ou menos acurado do que o modelo obtido por meio de análise discriminante linear. Essa análise discriminante tem o objeto de encontrar uma combinação que separam o grupo de dados em classes diferentes.

Na figura 9 é representado o fluxograma da metodologia deste trabalho, desde o banco de dados inserido, até a comparação final com o trabalho de Santos (2019).

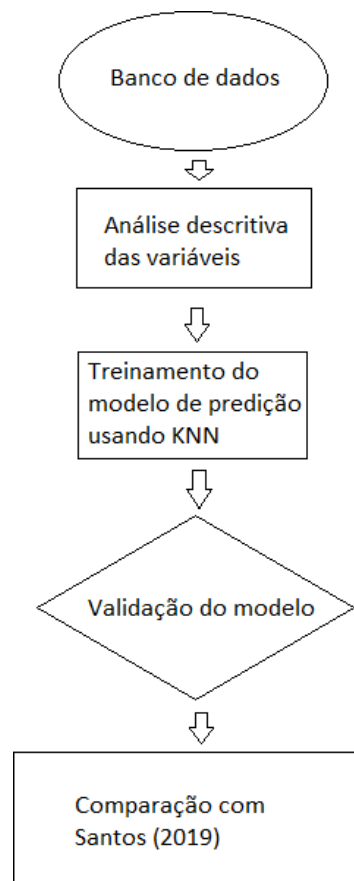


Figura 9 - Fluxograma do trabalho.

5. RESULTADOS E DISCUSSÕES

5.1. Análise descritiva do banco de dados

A partir dos dados obtidos do banco de dados, foram feitos levantamentos estatísticos sobre o que foi apresentados utilizando a média simples, o valor mínimo, valor máximo e o desvio padrão. A tabela 10 apresenta os dados de média, valores máximo e mínimos e desvio padrão das variáveis quantitativas do banco de dados.

	UCS (MPa)	Espaçamento (m)	Persistência da descontinuidade principal (m)	Abertura da descontinuidade principal (mm)	Altura do talude global (m)	Inclinação do talude global (°)
Média simples	77	2	9	2	321	42
Valor máximo	180	4	30	6,5	800	63
Valor mínimo	5	0,05	0,41	0,1	19	20
Desvio padrão	45	1	7	1	213	10

Tabela 10 - Análise descritiva das variáveis quantitativas do banco de dados.

Nota-se que para a resistência à compressão simples, os valores variam entre 5 e 180 MPa, sendo que, na natureza, esses valores podem variar de 1 MPa até mais de 250 MPa (ISRM, 1981). Sobre o espaçamento, o banco de dados possui maciços rochosos que apresentam variação entre 0,05 e 4 m, logo, elas variam de um grau de fraturamento maciço até um grau intensamente fraturado, não havendo assim rochas fragmentadas e desagregadas no banco. A persistência da descontinuidade principal dos maciços rochosos constituintes do banco de dados abrange desde as muito baixas (< 1 m), já que a menor do banco de dados tem 0,41 m, até as muito altas (> 20 m), uma vez que a maior do banco de dados tem 30 m. Quanto à abertura da descontinuidade principal, o banco de dados não apresenta descontinuidades muito fechadas ($< 0,1$ mm), já que a menor apresentada tem justamente 0,1 mm, o que ainda configura uma abertura fechada, e não apresenta aberturas cavernosas (> 1 m), já que a maior tem 6,5 mm.

Ao observar as alturas dos taludes globais, apesar de não terem uma classificação precisa para delimitar categorias, nota-se que os taludes do banco de dados pertencem a minas que estão em fases bem diferentes, uma vez que enquanto tem talude global com altura de 19 metros, provavelmente em fase inicial do projeto, tem talude global com altura de 800 metros, o que

implica em uma fase muito mais avançada. Essa grande variação também pode ser observada na inclinação dos taludes, uma vez que tem talude com inclinação de 20°, que é um valor de inclinação típico de taludes globais muito altos, até talude com 63°, que é um valor de inclinação mais comumente associado a bancadas e taludes mais baixos.

A rugosidade da descontinuidade principal é representada no gráfico da figura 10. Nota-se uma predominância de descontinuidades classificadas como rugosas e levemente rugosas, não existindo nenhuma classificada como polida e apenas uma classificada como muito rugosa.

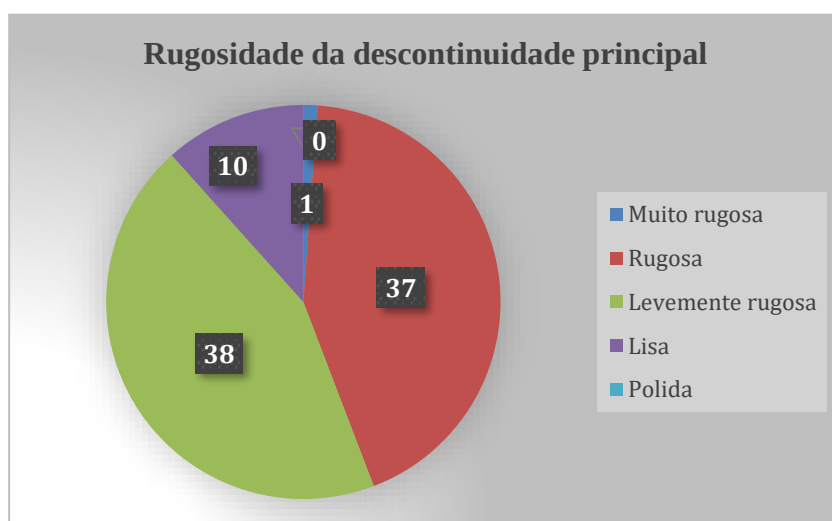


Figura 10 - gráfico de quantidade de taludes em cada categoria de rugosidade.

O preenchimento de descontinuidade principal também apresenta grandes diferenças na distribuição em cada uma das categorias estabelecidas, já que enquanto “Macio < 5 mm” e “Duro > 5 mm” englobam, respectivamente, 35 e 38 taludes, enquanto a categoria “nenhum” engloba três e “Macio > 5 mm” apenas um, como apresentado no gráfico da figura 11.

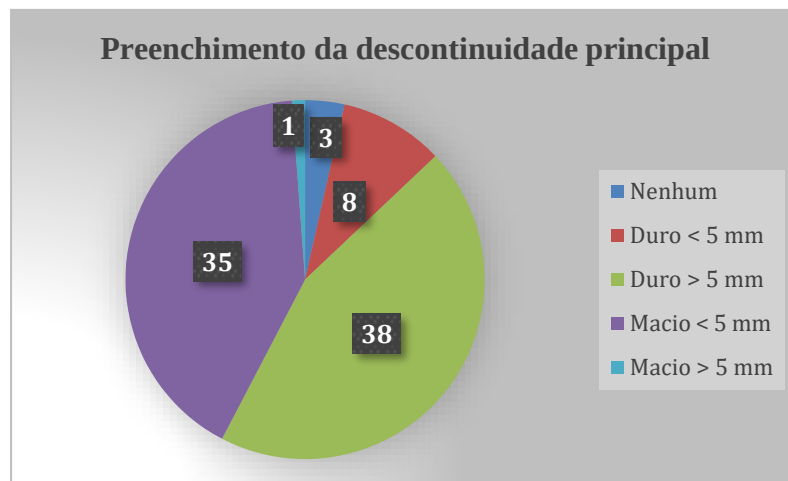


Figura 11 - Gráfico de quantidade de taludes em cada categoria de preenchimento.

O gráfico da figura 12 mostra que, sobre a alteração do maciço rochoso, a maior parte dos taludes no banco de dados apresenta rochas ligeiramente alteradas e moderadamente alteradas (34 e 33, respectivamente), enquanto existem poucos que apresentam rochas não alteradas (3) e decompostas (2).

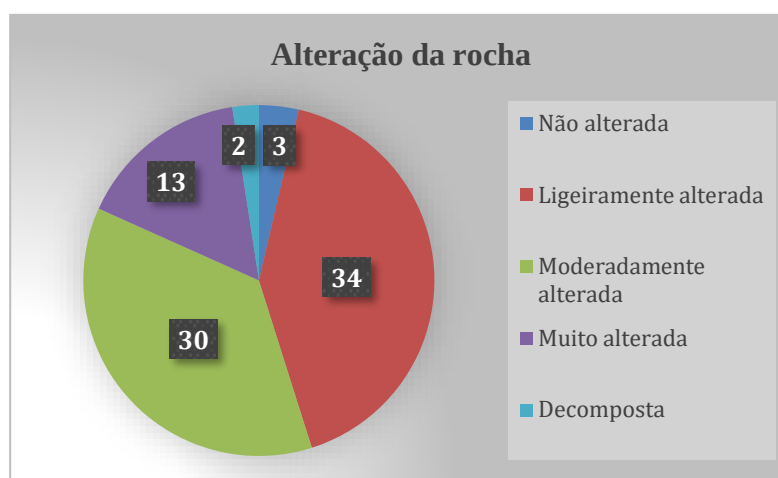


Figura 12 - Gráfico de quantidade de taludes em cada categoria de alteração da rocha.

Como pode se perceber no gráfico da figura 13, onde são representadas as categorias de condição de água subterrânea, que há um certo equilíbrio na quantidade de taludes nas categorias “encharcado”, “úmido” e “seco” (18, 30 e 22, respectivamente), mas são apenas 5 na categoria “gotejando” e nenhum “com fluxo”.

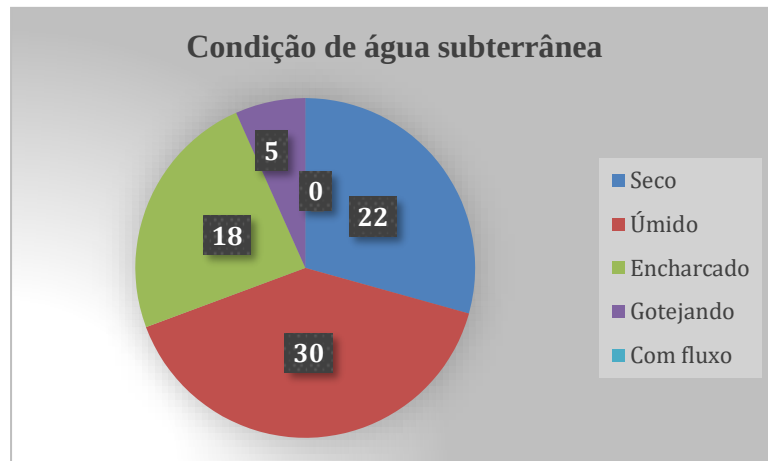


Figura 13 - Gráfico de quantidade de taludes em cada categoria de condição de água subterrânea.

Na orientação da descontinuidade principal, os quatro grupos presentes no banco de dados, como mostrado no gráfico da figura 14, apresentam um certo equilíbrio na quantidade de taludes englobados. Apesar disso, uma categoria não tem nenhum talude do banco de dados, que é a “muito desfavorável”.

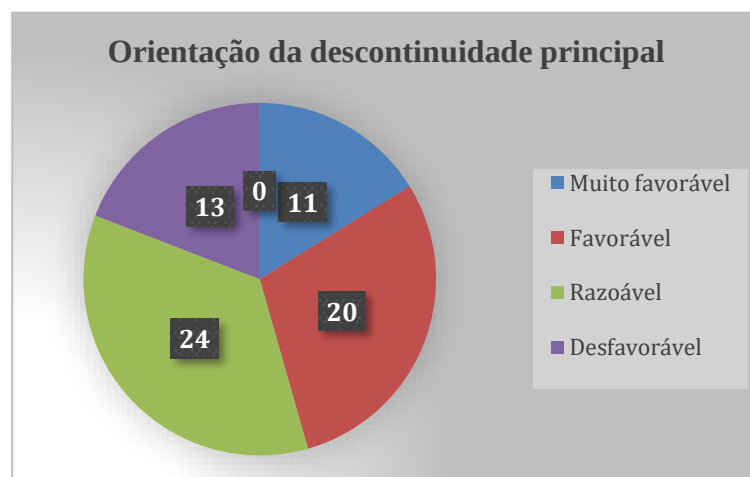


Figura 14 - Gráfico de quantidade de taludes em cada categoria de orientação da descontinuidade principal.

De acordo com o gráfico da figura 15, os métodos de desmonte mais comuns nos taludes que compõem o banco de dados são os de pré-fissuramento e o *modified production blast* com 24 taludes cada, enquanto o de desmonte regular tem apenas 4 e o de pós-fissuramento, nenhum.

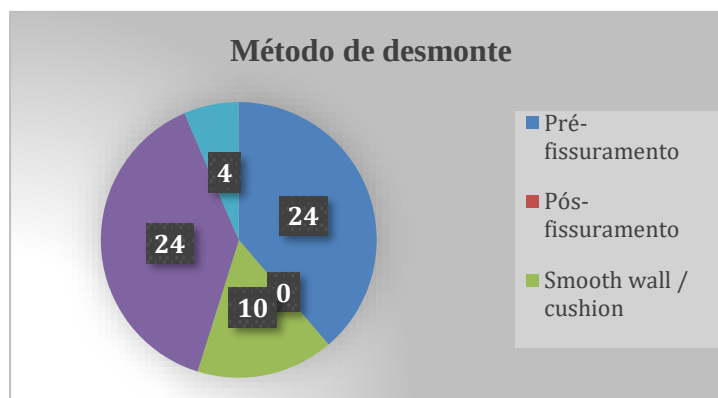


Figura 15 - Gráfico de quantidade de taludes em cada categoria de método de desmorte.

5.2. Modelo de predição e validação

Para treinamento e validação do modelo de predição da condição de estabilidade de taludes de mina, foi utilizado o programa R versão 4.2.0 (R Core Team, 2002). O algoritmo KNN foi aplicado ao banco de dados variando as proporções das amostras de treino e teste as composições de amostra de treino e teste (diferentes valores de semente) e os valores de K vizinhos. Os valores das amostras de treino foram variados segundo as seguintes proporções: 70% (30% de teste), 80% (20% de teste) e 90% (10% de teste). Quanto a definição das composições de amostra de treino e teste, os valores utilizados na fixação das sementes foram iguais a 000, 111, 222, 333, 444, 555, 666, 777, 888, 999 e 1000. Os valores de K utilizados variaram de 1 a 10.

Na tabela 11 são apresentadas as taxas globais de acerto do modelo de predição da condição de estabilidade de taludes de mina (acurácias), utilizando o banco de dados variando os valores das sementes e de K, considerando 70% de amostra de treino e 30% de amostra de teste. Na última linha da tabela foram inseridas as médias das acurácias para cada valor de K. Além disso, na figura 16, esses valores de acurácia são representados no gráfico onde cada linha representa um valor de semente.

	K									
Semente	1	2	3	4	5	6	7	8	9	10
000	55,56%	59,26%	70,37%	70,37%	77,78%	70,37%	66,67%	74,07%	66,67%	74,07%

111	74,07%	59,26%	66,67%	74,07%	66,67%	62,96%	62,96%	70,37%	70,37%	62,96%
222	74,07%	74,07%	81,48%	77,78%	77,78%	81,48%	85,19%	85,19%	85,19%	81,48%
333	85,19%	77,78%	77,78%	70,37%	74,07%	74,07%	77,78%	77,78%	66,67%	77,78%
444	66,67%	62,96%	74,07%	74,07%	70,37%	62,96%	66,67%	70,37%	66,67%	70,37%
555	55,56%	40,74%	66,67%	56,26%	62,96%	62,96%	77,78%	70,37%	77,78%	77,78%
666	77,78%	77,78%	66,67%	70,37%	70,37%	62,96%	62,96%	70,37%	66,67%	70,37%
777	66,67%	62,96%	66,67%	66,67%	62,96%	70,37%	66,67%	66,67%	70,37%	70,37%
888	66,67%	85,19%	74,07%	66,67%	59,26%	66,67%	62,96%	55,56%	62,96%	70,37%
999	74,07%	70,37%	81,48%	85,79%	81,48%	85,19%	77,78%	85,19%	74,07%	81,48%
1000	85,19%	55,56%	77,78%	81,48%	81,48%	77,78%	85,19%	81,48%	81,48%	81,48%
Média (K)	71,05%	65,99%	73,06%	72,17%	71,38%	70,71%	72,06%	73,40%	71,72%	74,41%

Tabela 11 - Acurácias do método variando as sementes e o K, considerando 70% de amostra de treino.

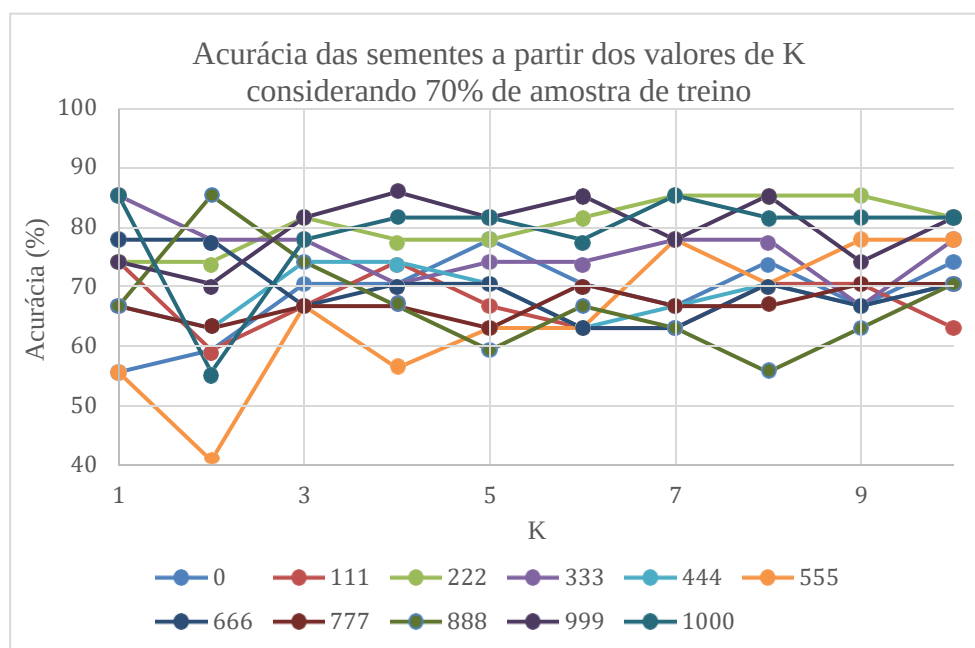


Figura 16 - Gráfico de acurácia das sementes a partir dos valores de K considerando 70% de amostra de treino.

A figura 17 ilustra o gráfico que, de acordo com os dados da tabela 11, mostra as médias das acurácias, quando utilizadas 70% das amostras de treino, para cada valor de K. No gráfico percebe-se que o valor de K que teve uma média mais alta de acurácia foi o K = 10, com 74,41%.

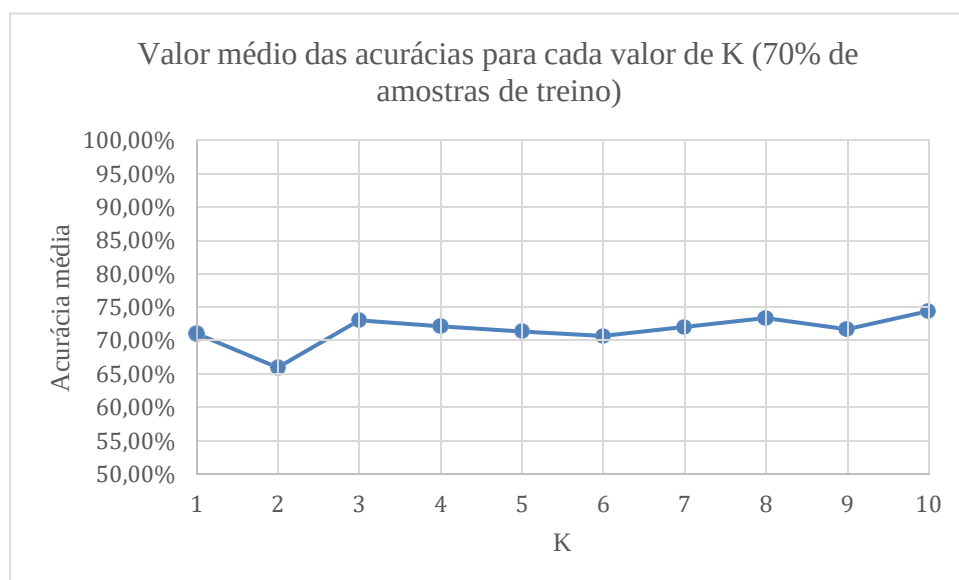


Figura 17 - Valor médio das acurácias para cada valor de K utilizando 70% das amostras de treino.

A tabela 12 representa os valores de acurácia do modelo de predição da condição de estabilidade de taludes de mina considerando 80% de amostra de treino e 20% de amostra de

teste, onde também pode se notar os valores das médias das acurácias para cada valor de K. O gráfico da figura 17, que ilustra os valores das acurácias a partir dos valores das sementes e de K, mostra valores ligeiramente mais elevados que os do gráfico da figura 16.

	K									
Semente	1	2	3	4	5	6	7	8	9	10
000	61,11%	66,67%	72,22%	72,22%	72,22%	72,22%	72,22%	72,22%	88,89%	66,67%
111	72,22%	61,11%	61,11%	83,33%	77,78%	66,67%	72,22%	77,78%	72,22%	77,78%
222	77,78%	72,22%	83,33%	88,89%	83,33%	77,78%	83,33%	88,89%	88,89%	88,89%
333	94,44%	66,67%	77,78%	77,78%	77,78%	77,78%	83,33%	72,22%	72,22%	72,22%
444	50,00%	66,67%	61,11%	66,67%	61,11%	61,11%	72,22%	55,56%	61,11%	61,11%
555	61,11%	50,00%	72,22%	77,78%	77,78%	83,33%	88,89%	88,89%	77,78%	83,33%
666	72,22%	77,78%	72,22%	72,22%	72,22%	72,22%	83,33%	83,33%	83,33%	77,78%
777	83,33%	66,67%	66,67%	77,78%	66,67%	77,78%	77,78%	77,78%	66,67%	77,78%
888	72,22%	77,78%	77,78%	77,78%	77,78%	72,22%	66,67%	72,22%	66,67%	72,22%
999	88,89%	77,78%	88,89%	83,33%	72,22%	83,33%	83,33%	88,89%	83,33%	77,78%
1000	83,33%	61,11%	77,78%	88,89%	77,78%	88,89%	88,89%	88,89%	88,89%	88,89%
Média (K)	74,24%	67,68%	73,74%	78,79%	74,24%	75,76%	79,29%	78,79%	77,27%	76,77%

Tabela 12 - Acurácias do método variando as sementes e o K, considerando 80% de amostras de treino.

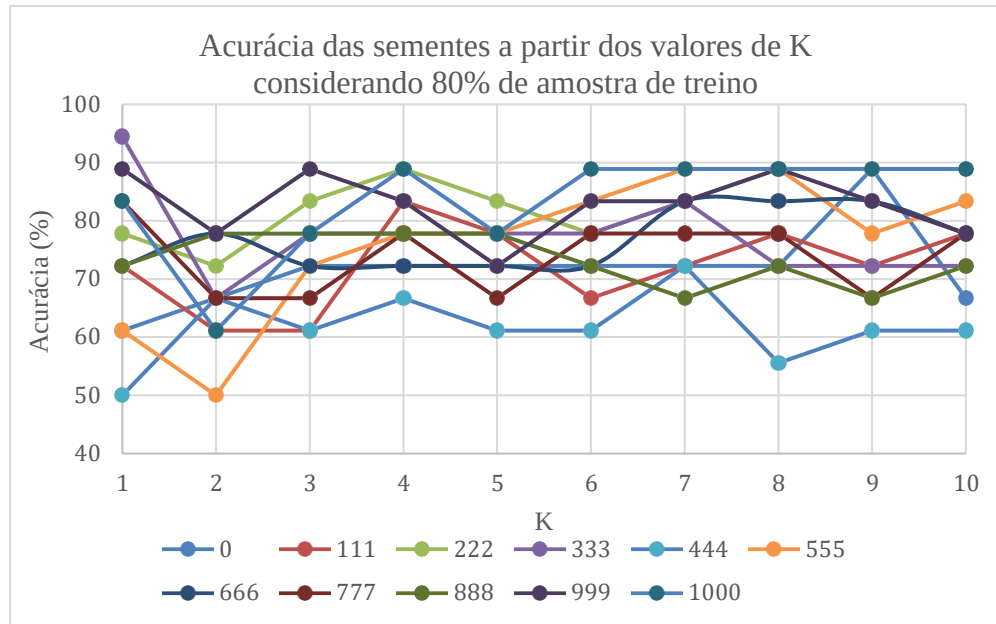


Figura 18 - Gráfico de acurácia das sementes a partir dos valores de K considerando 80% de amostra de treino.

A partir dos dados apresentados na última linha da tabela 12, a figura 19 ilustra o gráfico com as médias das acurácias para cada valor de K quando utilizadas 80% das amostras de treino e 20% de amostras de teste. Nota-se que o valor de K com maior média de acurácia foi K = 7, que obteve 79,29%.

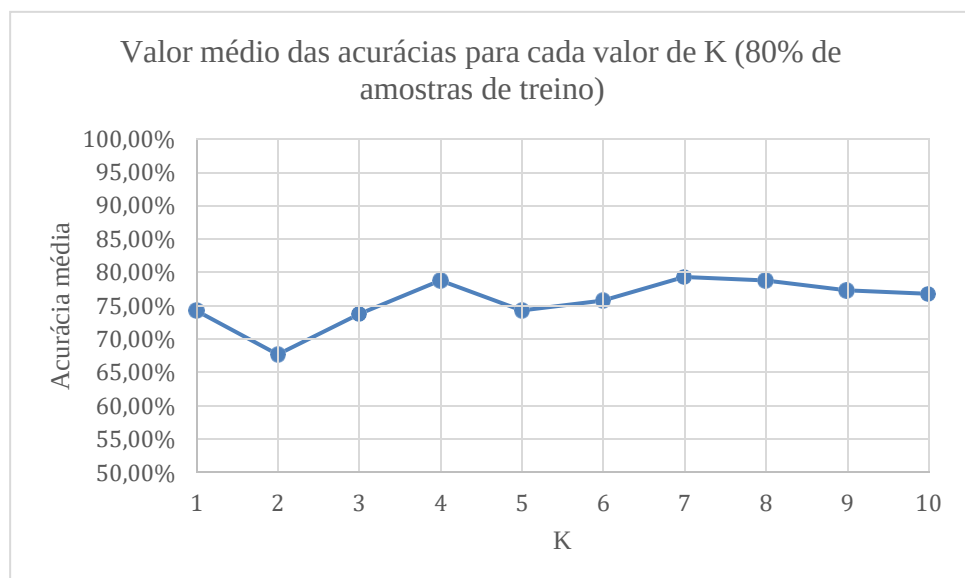


Figura 19 - Valor médio das acurácias para cada valor de K utilizando 80% das amostras de treino.

Enquanto isso, na tabela 13, onde foi utilizada 90% de amostras de treino e 10% de amostra de teste, é notório que são obtidos valores de acurácia ainda maiores que os apresentados nas tabelas 11 e 12, inclusive valores iguais a 100%. Esses valores maiores são ilustrados no gráfico da figura 18 onde nota-se muitos valores de acurácia superiores a 80%, o que justifica a média tão alta.

	K									
Semente	1	2	3	4	5	6	7	8	9	10
000	66,67%	55,56%	77,78%	77,78%	88,89%	88,89%	88,89%	88,89%	88,89%	88,89%
111	66,67%	55,56%	66,67%	77,78%	55,56%	88,89%	66,67%	77,78%	77,78%	88,89%
222	88,89%	77,78%	88,89%	77,78%	88,89%	88,89%	88,89%	88,89%	77,78%	77,78%
333	88,89%	77,78%	100,00%	88,89%	100,00%	100,00%	100,00%	88,89%	88,89%	100,00%
444	55,56%	55,56%	55,56%	66,67%	55,56%	44,44%	66,67%	77,78%	77,78%	77,78%
555	77,78%	66,67%	66,67%	77,78%	77,78%	88,89%	88,89%	88,89%	88,89%	77,78%
666	66,67%	66,67%	77,78%	77,78%	77,78%	77,78%	88,89%	88,89%	88,89%	100,00%
777	88,89%	55,56%	66,67%	66,67%	77,78%	66,67%	77,78%	66,67%	77,78%	66,67%
888	77,78%	77,78%	66,67%	66,67%	66,67%	66,67%	77,78%	66,67%	77,78%	77,78%
999	88,89%	77,78%	66,67%	66,67%	77,78%	88,89%	88,89%	88,89%	77,78%	66,67%
1000	77,78%	66,67%	77,78%	88,89%	77,78%	88,89%	88,89%	88,89%	88,89%	88,89%
Média (K)	76,77%	66,67%	73,74%	75,76%	76,77%	80,81%	83,84%	82,83%	82,83%	82,83%

Tabela 13 - Acurácias do método variando as sementes e o K, considerando 90%.

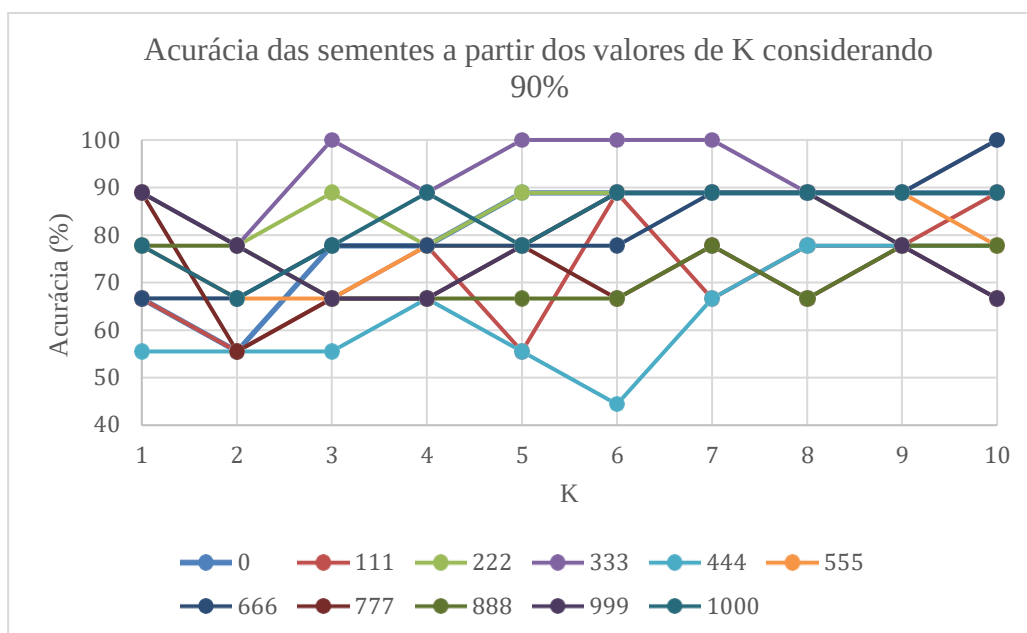


Figura 20 - Gráfico de acurácia das sementes a partir dos valores de K considerando 90% de amostra de treino.

A figura 21 ilustra o gráfico das médias das acurácias para cada valor de K considerando 90% das amostras de treino e 10% de amostra de teste, como apresentado na tabela 13. Nele observa-se que a quantidade de vizinhos que mais aproximou de uma acurácia ideal foi K = 7, com o valor médio de 83,84%.

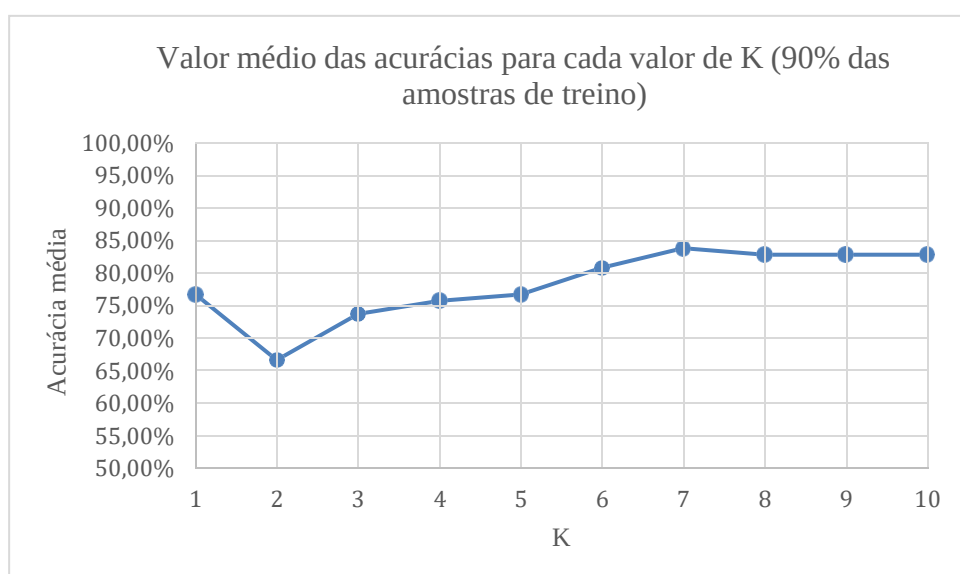


Figura 21 - Valor médio das acurácias para cada valor de K utilizando 90% das amostras de treino.

5.3. Comparação com o modelo obtido por Santos (2019)

Para que seja feita a comparação entre os métodos utilizados nos dois trabalhos, é fundamental que seja levado em conta que quando Santos (2019) desenvolveu o modelo de análise discriminante, o método de validação utilizado foi o método da ressubstituição, em que as amostras utilizadas para o treinamento do modelo são utilizadas para testar o modelo também. Não foi estabelecida uma configuração de divisão entre amostra de treino teste e não foram feitas as variações no número de sementes e, portanto, o total dos 88 taludes foram utilizados tanto no treinamento como no teste. Como consequência disso, o resultado de acurácia obtido pelo modelo de Santos (2019) tende a ser superestimado.

No modelo desenvolvido por Santos (2019), o erro obtido foi de 11,36% enquanto no modelo deste trabalho a maior acurácia, foi encontrada para uma composição de 90% de amostras para treino, 10% de amostra para teste e para K igual a 7. A acurácia obtida foi igual a 83,84%, logo, o erro é de 16,16%. Apesar do erro obtido por Santos (2019) ser menor, devido ao fato da metodologia de avaliação dos métodos ter sido feito de maneira diferente, como já citado, a comparação direta dos resultados deve ser realizada com cuidado.

Como exemplo dessa diferença de metodologias para avaliação da acurácia do método, caso fosse analisada apenas a divisão de 90% de amostra de treino com valor de semente igual a 333, como pode ser visto na tabela 13, a acurácia seria 100%, ou seja, o erro do modelo seria nulo, da mesma forma que caso sejam observadas apenas uma das outras divisões com alguns outros valores de semente, o erro seria muito maior que o encontrado. Isso mostra que a análise da acurácia deve ser feita da forma abrangente, levando diferentes divisões de amostra de treinamento e de teste para que seja obtido um resultado preciso.

6. CONCLUSÃO

Diante de todo o trabalho realizado, nota-se que é fundamental o uso de um banco de dados robusto e que compreenda a maior quantidade de variáveis e amostras possível, uma vez que a estabilidade de um talude depende de uma série de fatores e a análise minuciosa deles trará resultados mais precisos no estudo de novos casos. O banco de dados utilizado possui diversos taludes que estão representados em quase todas as categorias possíveis das variáveis estudadas, aumentando a eficiência do trabalho.

A partir do banco de dados foi utilizado o software R para a implementação do método KNN que variava a porcentagem das amostras de teste do banco de dados, configuração das amostras de treino e teste (sementes) e o valor de K, ou seja, os vizinhos mais próximos utilizados para a análise. Sendo assim, foi gerada uma matriz de confusão de onde foi calculada a acurácia do modelo de predição da condição de estabilidade de taludes para diferentes configurações de parâmetros.

Quando analisados os resultados obtidos variando cada uma das porcentagens das amostras de treino, das sementes e dos valores de K, é possível notar que quanto maior a porcentagem das amostras de teste, maior as acurácias obtidas variando as sementes e os valores de K, como é evidenciado pelas médias gerais das acurácias quando considerado 70% das amostras de teste (71,59% de acurácia média), 80% das amostras de teste (75,66% de acurácia média) e 90% das amostras de teste (78,28% de acurácia média).

A acurácia obtida neste trabalho foi menor do que a obtida por Santos (2019), quando foi utilizada análise discriminante e isso se deve ao fato de as metodologias de avaliação das acurácias terem sido diferentes. No trabalho desenvolvido por Santos (2019) não foram utilizados diferentes valores de sementes e nem foram variados os valores das amostras de treino e teste, o que pode gerar uma acurácia alta e imprecisa, já que a análise se limita a apenas um valor de semente e os mesmos dados utilizados para treino, foram usados para teste do método, justificando a superioridade do método de Santos (2019).

Todos os resultados obtidos durante o trabalho são baseados nas variáveis utilizadas e nos 88 taludes do banco de dados. Tanto as entradas, como porcentagem da amostra de treino, semente e valor de K, podem ser alteradas, como também pode ser inserido um banco de dados maior assim como a quantidade de variáveis de estudo, para uma análise mais detalhada e resultados mais eficientes.

REFERÊNCIAS BIBLIOGRÁFICAS

- AZANK, F. **Modelos de Predição | KNN**. Turing Talks. Disponível em <<https://medium.com/turing-talks/turing-talks-13-modelo-de-predi%C3%A7%C3%A3o-knn-3be880c9b9d1>>. 2019.
- BARTON, N. R & Choubey, V. **The shear strength of rock joints in theory and practice**. Rock Mechanics and Rock Engineering, ResearchGate. 1977.
- BIENIAWSKI, Z. T. **Engineering Rock Mass Classifications**. New York: John Wiley & Sons. 1989.
- BRANCO, P. M. **Rochas**. Disponível em <<http://www.cprm.gov.br/publique/SGB-Divulga/Canal-Escola/Rochas-1107.html#:~:text=Rocha%20%C3%A9%20uma%20associa%C3%A7%C3%A3o%20natural,%2C%20muito%20frequentemente%2C%20tamb%C3%A9m%20mica>> Serviço Geológico do Brasil - CPRM. 2015.
- BUI, D. T. *et al.* **Predicting Slope Stability Failure through Machine Learning Paradigms**. MDPI, International Journal of Geo-Information. 2019.
- DEERE, D. U. **Rock Quality Designation (RQD) After Twenty Years**. National Technical Information Service. Gainesville, Florida. p 22. 1989.
- FILHO, C. L. M. & NUMMER, A. V. **Introdução à Geologia de Engenharia**. 5ª edição - Revista e Ampliada, editora ufsm. p 4.1 a 4.91. 2014.
- FIORI, A. P. & CARMIGNANI, L. **Fundamentos de mecânica dos solos e das rochas: aplicações na estabilidade de taludes**. 2ª edição, Revisão de Patrícia Domingues Ribas. 2013.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Massachusetts Institute of Technology. p 98 a 155. 2016.
- HOEK, E. & MARINOS, P. **Predicting Tunnel Squeezing Problems in Weak Heterogeneous Rock Masses**. Tunnels and Tunnelling. 2000.

HUANG, S. *et al.* **An Improved KNN-Based Slope Stability Prediction Model**. Hindawi, Advances in Civil Engineering, Tayfun Dede. 2020.

ISRM. **Suggested Methods for the Quantitative Description of Discontinuities in Rock Mass**. Grã Bretanha, v. 15. p 319 - 368. 1978.

JAIQUES, D. S. **Caracterização e Classificação de Maciços Rochosos da Mina de Volta Grande, Nazareno, Minas Gerais**. Dissertação de Mestrado, Universidade Federal de Viçosa, Programa de Pós-Graduação em Engenharia Civil. Viçosa/MG, 2014.

LACERDA, S.G. *et al.* **Caracterização geomecânica do maciço rochoso da Gruta do Pião - Parque Estadual do Ibitipoca, Minas Gerais**. In: ZAMPAULO, R. A. (org.) CONGRESSO BRASILEIRO DE ESPELEOLOGIA, 35, 2019. Bonito. Anais... Campinas: SBE, 2019. p.244-252. Disponível em:

MARIZ, F. M. **AVALIAÇÃO E COMPARAÇÃO DE VERSÕES MODIFICADAS DO ALGORITMO KNN**. Trabalho de Conclusão de Curso no curso de Ciência da Computação no Centro de Informática da Universidade Federal de Pernambuco. Recife, 2017.

MESQUITA, J. B. **Maciços Rochosos em Projetos Hidroelétricos: Proposta de Classificação Geomecânica**. Dissertação de Doutorado, Universidade Estadual Paulista, Instituto de Geociências e Ciências Exatas. Rio Claro/SP, 2008.

MOAYEDI, H. *et al.* **Machine-Learning-Based Classification Approaches toward Recognizing Slope Stability Failure**. MDPI, Applied Sciences. 2019.

NAGHADEHI, Z. *et al.* **A new open-pit mine slope instability index defined using the improved rock engineering systems approach**. International Journal of Rock Mechanics & Mining Sciences. 2013.

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.

PASSO, D. P. **ANÁLISE DA QUALIDADE DE CLASSIFICADORES PARA IDENTIFICAÇÃO DE ALVOS URBANOS EM IMAGENS DE ALTA RESOLUÇÃO ESPACIAL - UMA APLICAÇÃO COM AS IMAGENS DO SATÉLITE WORLDVIEW II**. Dissertação de mestrado apresentada no Instituto de Geociências Aplicadas da Universidade de Brasília. Brasília-DF, 2013.

PHAM, K. **Ensemble learning-based classification models for slope stability analysis**. Elsevier, Catena. 2021.

READ, J & STACEY, P. **GUIDELINES FOR OPEN PIT SLOPE DESIGN**. Csiro Publishing. p 206 a 211. 2010.

REIS, E. A. & REAIS, I. A. **Análise Descritiva dos dados**. Relatório Técnico do Departamento de Estatística da UFMG. 1ª edição, 2002.

SANTOS, A. E. M. **CLASSIFICAÇÃO DE MACIÇOS ROCHOSOS POR MEIO DE TÉCNICAS DA ESTATÍSTICA MULTIVARIADA E INTELIGÊNCIA ARTIFICIAL**. Tese do programa de Pós-Graduação em Engenharia Mineral do Departamento de Engenharia de Minas da Universidade Federal de Ouro Preto para Mestrado. Ouro Preto/MG, 2021.

SANTOS, T. B. **ANÁLISE DE RISCO GEOTÉCNICO EM TALUDES ROCHOSOS DE MINA COM USO DE TECNOLOGIAS ESTATÍSTICAS MULTIVARIADAS E DE APRENDIZADO DE MÁQUINA**. Tese de doutorado em Engenharia Mineral apresentada ao Programa de Pós-Graduação do Departamento de Engenharia de Minas da Escola de Minas da Universidade Federal de Ouro Preto. Ouro Preto/MG, 2019.

SANTOS, T. B. *et al.* **Quantitative hazard assessment system (Has-Q) for open pit mine slopes**. International Journal of Mining Science and Technology. 2019.

SILVA, L. M. O. **UMA APLICAÇÃO DE ÁRVORE DE DECISÃO, REDES NEURAIS E KNN PARA A IDENTIFICAÇÃO DE MODELOS ARMA NÃO-SAZONAIS E SAZONAIS**. Programa PPG em Engenharia Elétrica para título de Doutor em Engenharia Elétrica pela Pontifícia Universidade Católica do Rio de Janeiro. Rio de Janeiro, 2005.

SJOBERG, J. LARGE SCALE SLOPE STABILITY IN OPEN PIT MINING - A REVIEW. Relatório técnico 1996:10T, Divisão de Mecânica das Rochas da Universidade de Tecnologia de Lulea. Suécia, 1996.

WOLPP, L. F. L. CARACTERIZAÇÃO E CLASSIFICAÇÃO GEOMECÂNICA DE UM TALUDE DA SANTA HELENA DE MINERAÇÃO LTDA. Trabalho de conclusão de curso, Engenharia de minas, Centro Federal de Educação Tecnológica de Minas Gerais. Araxá/MG, 2018.

WYLLIE, D. C. & MAH, C. W. ROCK SLOPE ENGINEERING CIVIL AND MINING. 4ª edição. 2004.

ANEXO

ANEXO: Banco de dados utilizado no trabalho

Esp.	Persist. da desc. principal	Abertura da desc. principal	Rugosidade da desc. principal	Pre. da desc. principal	Alt. da rocha	Condição de água subterrânea	Orientação da descontin. principal	Mét. de desm.	Alt. do tal. Glob. (m)	Inclin. do talude global (°)	Status
0,75	12,5	3	2	1	2	5	4	1	170	30	OF
0,75	25	3	3	2	3	5	4	1	170	20	ST
1,50	22,5	1	3	2	3	5	3	1	130	21	ST
1,50	7,5	1	4	2	4	5	3	1	120	27	ST
1,00	30	3	3	2	4	5	4	1	210	55	ST
0,75	22,5	3	4	4	3	5	1	1	195	35	ST
2,00	10	1,05	3	3	5	5	4	1	210	55	ST
2,00	10	1,05	3	3	5	5	4	1	210	55	ST
1,00	25	3	3	2	4	5	4	1	210	45	ST
1,20	22,5	1	4	3	4	5	3	1	190	47	ST
1,10	12,5	3	4	3	2	4	3	1	190	40	FSB
1,25	25	6,5	3	3	2	4	5	1	185	40	OF
1,30	12,5	3	4	3	3	4	4	1	230	40	FSB
1,25	7,5	3	3	3	3	5	4	1	175	45	ST
1,25	2,25	1,25	3	3	4	5	3	1	175	45	ST
0,18	1,6	0,75	4	2	3	5	2	1	160	38	ST
0,15	1,6	1,5	4	2	3	5	3	1	175	45	ST
0,22	2,25	1,25	3	3	4	5	3	1	175	45	ST
1,50	6,5	3	4	3	4	5	2	2	420	35	ST
2,50	17,5	3	3	1	4	4	3	2	620	35	FSB
3,00	17,5	3	3	2	4	4	4	2	800	35	FSB
4,00	12,5	3	4	3	4	5	4	2	620	35	ST
1,30	15	3	3	2	3	5	2	1	460	35	ST
2,00	6,5	6	4	2	4	4	3	1	280	35	ST
1,20	15	3	4	2	2	4	3	1	490	35	FSB
1,20	12,5	3	4	3	3	4	3	1	340	35	ST
0,82	0,91	3	3	3	3	3	5	5	96	56	OF
0,31	0,5	3	4	4	3	4	2	5	180	55	ST
0,49	1,36	3	4	2	2	3	4	5	180	42	FSB
0,55	0,64	3	3	3	2	2	5	5	180	48	OF
0,66	0,41	0,55	3	3	3	3	4	5	180	56	FSB
0,14	0,66	0,55	4	3	3	4	3	5	96	50	ST
0,36	1,06	3	3	4	4	4	3	5	96	56	ST
2,50	2,5	3	3	3	2	3	3	2	312	57	FSB
2,00	6,5	3	3	3	2	3	4	2	235	37	OF
2,00	6,75	3	4	4	3	4	3	2	234	36	ST

2,50	7,25	3	2	2	2	2	4	2	20	48	OF
2,15	2,5	0,55	3	3	2	3	3	2	288	63	FSB
4,00	5,75	0,55	3	3	4	5	3	5	750	31	ST
2,00	4,5	0,55	2	3	3	5	4	5	750	32	ST
3,30	5,75	3	3	2	3	4	5	5	750	32	OF
3,50	2,75	0,55	4	3	3	5	3	5	700	31	ST
2,40	7,75	3	4	2	4	5	5	5	780	42	FSB
1,00	22,5	3	3	3	4	3	5	5	325	58	FSB
0,90	27,5	3	3	2	3	2	5	5	325	58	FSB
0,60	12,5	3	2	2	3	3	4	5	325	58	FSB
1,65	22,5	3	3	2	4	3	4	5	325	51	FSB
3,00	1,75	0,55	4	4	4	5	2	5	325	51	ST
0,70	22,5	3	4	3	3	4	3	5	325	51	FSB
3,20	6,5	0,55	4	2	4	5	3	5	500	58	ST
1,75	3	3	5	2	4	4	2	5	500	58	ST
1,00	4,5	0,55	4	2	4	5	4	5	500	46	ST
0,80	5,5	3	3	2	3	3	4	5	500	52	FSB
3,70	4,5	3	4	2	4	5	3	5	500	58	ST
1,50	9,5	3	3	3	3	4	4	3	230	47	FSB
2,70	3	0,55	4	2	5	5	3	3	250	51	ST
2,70	4,75	3	4	2	4	4	4	3	145	56	FSB
1,50	3,5	0,55	4	2	4	5	3	3	325	53	ST
2,70	2,5	0,1	4	2	4	5	2	3	325	49	ST
3,50	2,25	0,55	3	3	4	4	2	3	125	42	ST
1,30	6,5	3	2	3	3	3	5	2	700	42	OF
1,25	6,5	0,55	4	3	3	4	3	2	670	40	FSB
1,50	3,5	0,55	4	3	4	4	2	2	670	40	ST
3,00	3,5	0,55	4	3	3	4	2	2	710	40	ST
3,00	3,5	0,55	3	4	4	4	3	2	725	40	ST
0,60	6,5	0,55	2	2	3	3	5	2	740	32	FSB
2,00	7,5	3	3	2	3	3	5	2	690	35	OF
0,60	7,5	0,55	3	3	4	4	3	2	240	32	FSB
2,25	3	0,1	4	3	4	5	2	2	240	34	ST
0,80	10	3	3	2	3	3	4	2	210	34	OF
0,80	10	3	3	3	4	3	5	2	210	34	OF
0,80	15	3	2	2	3	4	4	2	210	32	OF
1,30	17,5	3	4	3	2	2	3	2	280	45	OF
1,00	17,5	3	3	4	3	3	3	2	440	28	OF
0,70	17,5	5	3	4	2	2	5	2	280	35	OF
1,30	17,5	5	3	2	2	3	4	2	550	38	OF
2,00	4	0,55	4	3	4	4	2	2	260	26	ST

1,00	20	3	3	2	4	4	4	3	100	40	OF
1,40	4	3	4	2	4	5	3	3	100	40	ST
2,70	6,5	3	4	2	3	5	4	2	150	45	FSB
3,00	7,5	0,55	4	5	4	4	3	5	300	50	ST
0,90	10	5	4	3	3	4	3	2	305	50	FSB
0,60	5,25	3	3	3	4	4	2	3	450	28	FSB
0,50	5,25	3	3	3	3	3	4	3	250	35	OF
0,2	6	0,5	2	5	1	4	3	1	50	45	FSB
0,05	12	0,3	2	2	4	4	2	1	19	52	FSB
0,05	10	2	2	5	2	4	2	1	19	51	FSB
0,85	11	0,5	2	2	1	4	4	1	50	48	OF

Tabela 14 - Banco de dados utilizado no trabalho.

APÊNDICE

APENDICE: Script do software R utilizado para a implementação do método KNN utilizando 90% das amostras de teste, semente = 1000 e K = 7.

```
## Leitura dos dados
```

```
dados <- read.table("taludes.txt",header=TRUE, row.names = 1, dec = ',')
```

```
dados=as.data.frame(cbind(dados))
```

```
View(dados)
```

```
##Generate a random number that is 90% of the total number of rows in dataset.
```

```
set.seed(1000)
```

```
ran <- sample(1:nrow(dados), 0.90 * nrow(dados))
```

```
ran
```

```
##stadardization
```

```
dados_norm<- as.data.frame(scale(dados[,c(1:12)]))
```

```
##extract training set
```

```

dados_train <- dados_norm[ran,]

dados_train

##extract testing set

dados_test <- dados_norm[-ran,]

View(dados_test)

##extract 13th column of train dataset because it will be used as 'cl' argument in knn function.

dados_target_category <- dados[ran,13]

dados_target_category

##extract 13th column if test dataset to measure the accuracy

dados_test_category <- dados[-ran,13]

dados_test_category

##load the package class

library(class)

##run knn function

pr <- knn(dados_train,dados_test,cl= dados_target_category,k=7)

##create confusion matrix

tab <- table(pr,dados_test_category)

tab

```

##this function divides the correct predictions by total number of predictions that tell us how accurate teh model is.

```
accuracy <- function(x){sum(diag(x)/(sum(rowSums(x)))) * 100}
```

```
accuracy(tab)
```