



Universidade Federal de Ouro Preto  
Escola de Minas  
CECAU - Colegiado do Curso de  
Engenharia de Controle e Automação



Rodolfo César de Oliveira Pereira

**Uso de *machine learning* na predição do risco de movimentos de  
massa no município de Ouro Preto - MG**

Monografia de Graduação

Ouro Preto, 2024

Rodolfo César de Oliveira Pereira

**Uso de *machine learning* na predição do risco de movimentos de massa no município de Ouro Preto - MG**

Trabalho apresentado ao Colegiado do Curso de Engenharia de Controle e Automação da Universidade Federal de Ouro Preto como parte dos requisitos para a obtenção do Grau de Engenheiro(a) de Controle e Automação.

Universidade Federal de Ouro Preto

Orientador: Prof. Dr. Bruno N. Coelho

Ouro Preto

2024



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



**FOLHA DE APROVAÇÃO**

**Rodolfo César de Oliveira Pereira**

**Uso de *machine learning* na predição do risco de movimentos de massa no município de Ouro Preto -  
MG**

Monografia apresentada ao Curso de Engenharia de Controle e Automação da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Engenheiro de Controle e Automação

Aprovada em 23 de fevereiro de 2024

Membros da banca

[Dr.] - Bruno Nazário Coelho - Orientador (Universidade Federal de Ouro Preto)  
Frederico Luiz Martins de Sousa (PROFICAM - Universidade Federal de Ouro Preto)  
[Dr.] - Daniel Silva Jaques- (Universidade Federal de Ouro Preto)

Bruno Nazário Coelho, orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 27/02/2024



Documento assinado eletronicamente por **Bruno Nazário Coelho, PROFESSOR DE MAGISTERIO SUPERIOR**, em 27/02/2024, às 23:24, 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 **0672579** e o código CRC **3DC7619C**.

**Referência:** Caso responda este documento, indicar expressamente o Processo nº 23109.002107/2024-65

SEI nº 0672579

R. Diogo de Vasconcelos, 122, - Bairro Pilar Ouro Preto/MG, CEP 35402-163  
Telefone: 3135591533 - [www.ufop.br](http://www.ufop.br)

# Agradecimentos

Primeiramente a Deus, aos meus pais e irmão, por me proporcionarem a oportunidade de me formar como engenheiro de controle e automação e me apoiarem durante todo este percurso.

Aos integrantes da Defesa Civil de Ouro Preto, pela disponibilidade em me ceder os dados e por contribuírem com discussões fundamentais ao meu estudo.

Ao meu orientador, que de prontidão apoiou a minha ideia de trabalho e me ajudou a torná-la realidade.

Aos integrantes da banca avaliadora, pelas pertinentes contribuições ao meu trabalho.

Aos meus colegas de curso, pelas discussões e trocas de conhecimentos ao longo destes últimos cinco anos.

E a todos que de alguma forma contribuíram com a realização deste meu objetivo.

*A vida é um ininterrupto vir a ser, jamais um ser puro e causal.*

---

— Albert Einstein.

# Resumo

Ouro Preto é o município brasileiro com o maior número de áreas de risco geológico, 313, o que afeta diretamente a vida de mais de três mil pessoas que habitam estas áreas. Tendo em vista este cenário, o presente estudo visa avaliar a capacidade preditiva de algoritmos de *machine learning* quanto aos eventos de movimentos de massa na cidade de Ouro Preto-MG. Foram utilizados dados pluviométricos e registros de ocorrências de movimentos de massa da Defesa Civil no período compreendido entre 2012 a 2022. Os algoritmos apresentaram melhores resultados quando os dados utilizados foram balanceados, sendo o melhor desempenho da Floresta Aleatória com acurácia de 0,74 e *recall*, para o evento de interesse, de 0,83. Contudo, os resultados obtidos podem ser aprimorados em estudos futuros com a utilização de uma base de dados mais sólida e de novos parâmetros de entrada para os algoritmos.

**Palavras-chaves:** Aprendizado de máquina. Movimentos de massa. Floresta aleatória.

# Abstract

Ouro Preto is the Brazilian municipality with the largest number of geological risk areas, 313, which directly affects the lives of more than three thousand people who live in these areas. Given this scenario, the present study aims to evaluate the predictive capacity of *machine learning* algorithms regarding mass movement events in the city of Ouro Preto-MG. Rainfall data and records of mass wasting occurrences from Civil Defense were used in the period between 2012 and 2022. The algorithms presented better results when the data used were balanced, with the best performance being the Random Forest with an accuracy of 0.74 and *emph*recall, for the event of interest, of 0.83. However, the results obtained can be improved in future studies with the use of a more solid database and new input parameters for the algorithms.

**Key-words:** *Machine learning*. Mass wasting. Random forest.

# Lista de ilustrações

Figura 1 – Árvore de decisão . . . . .	15
Figura 2 – Floresta aleatória . . . . .	16
Figura 3 – Curva logística . . . . .	17
Figura 4 – <i>Support Vector Machine</i> . . . . .	17
Figura 5 – Gráfico do acumulado pluviométrico anual . . . . .	21
Figura 6 – Mapa de declividade do município . . . . .	22
Figura 7 – Mapa de susceptibilidade a escorregamentos no município . . . . .	23
Figura 8 – Pré-processamento dos dados . . . . .	24
Figura 9 – Distribuição dos dados brutos . . . . .	26
Figura 10 – Matriz de confusão - Árvore de Decisão com dados brutos . . . . .	27
Figura 11 – Matriz de confusão - Floresta Aleatória com dados brutos . . . . .	27
Figura 12 – Matriz de confusão - Regressão Logística com dados brutos . . . . .	28
Figura 13 – Matriz de confusão - SVM com dados brutos . . . . .	29
Figura 14 – Distribuição dos dados balanceados . . . . .	30
Figura 15 – Matriz de confusão - Árvore de Decisão com dados balanceados . . . . .	31
Figura 16 – Matriz de confusão - Floresta Aleatória com dados balanceados . . . . .	31
Figura 17 – Matriz de confusão - Regressão Logística com dados balanceados . . . . .	32
Figura 18 – Matriz de confusão - SVM com dados balanceados . . . . .	33

# Lista de tabelas

Tabela 1 – Tabela das métricas de desempenho - Árvore de Decisão com dados brutos . . . . .	27
Tabela 2 – Tabela das métricas de desempenho - Floresta Aleatória com dados brutos . . . . .	28
Tabela 3 – Tabela das métricas de desempenho - Regressão Logística com dados brutos . . . . .	28
Tabela 4 – Tabela das métricas de desempenho - SVM com dados brutos . . . . .	29
Tabela 5 – Tabela das métricas de desempenho - Árvore de Decisão com dados balanceados . . . . .	31
Tabela 6 – Tabela das métricas de desempenho - Floresta Aleatória com dados balanceados . . . . .	32
Tabela 7 – Tabela das métricas de desempenho - Regressão Logística com dados balanceados . . . . .	32
Tabela 8 – Tabela das métricas de desempenho - SVM com dados balanceados . . . . .	33
Tabela 9 – Tabela de análise dos dados falsos-positivos do algoritmo floresta aleatória	34

# Lista de abreviaturas e siglas

Cemaden	Centro Nacional de Monitoramento e Alertas de Desastres Naturais
CPRM	Companhia de Pesquisa de Recursos Minerais
Cwb	<i>Temperate/Dry winter/Warm summer</i>
FN	Falso Negativo
FP	Falso Positivo
PROMOSAT-OP	Programa de Monitoramento Socioambiental do Território Ouro-pretano
SVM	<i>Support Vector Machine</i>
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>12</b>
<b>1.1</b>	<b>Justificativas e Relevância</b>	<b>12</b>
<b>1.2</b>	<b>Objetivos</b>	<b>12</b>
<b>2</b>	<b>REVISÃO DE LITERATURA</b>	<b>14</b>
<b>2.1</b>	<b><i>Machine Learning</i></b>	<b>14</b>
2.1.1	Árvore de Decisão	14
2.1.2	Floresta Aleatória	15
2.1.3	Regressão Logística	16
2.1.4	<i>Support Vector Machine</i>	17
2.1.5	Métricas de desempenho	18
2.1.5.1	Matriz de confusão	18
2.1.5.2	Acurácia	18
2.1.5.3	Precisão	19
2.1.5.4	<i>Recall</i>	19
2.1.5.5	<i>F1-score</i>	19
<b>2.2</b>	<b>Movimentos de massa</b>	<b>19</b>
<b>2.3</b>	<b>Local de estudo</b>	<b>20</b>
2.3.1	Localização geográfica	20
2.3.2	Climatologia	21
2.3.3	Topografia e declividade	22
2.3.4	Áreas de risco	22
<b>3</b>	<b>MATERIAIS E MÉTODOS</b>	<b>24</b>
<b>3.1</b>	<b>Coleta dos dados</b>	<b>24</b>
<b>3.2</b>	<b>Pré-processamento dos dados</b>	<b>24</b>
<b>3.3</b>	<b>Algoritmos</b>	<b>25</b>
<b>3.4</b>	<b>Métricas de desempenho</b>	<b>25</b>
<b>4</b>	<b>RESULTADOS E DISCUSSÃO</b>	<b>26</b>
<b>4.1</b>	<b>Dados brutos</b>	<b>26</b>
4.1.1	Árvore de decisão	26
4.1.2	Floresta aleatória	27
4.1.3	Regressão logística	28
4.1.4	<i>Support Vector Machine</i>	29
<b>4.2</b>	<b>Dados balanceados</b>	<b>30</b>

4.2.1	Árvore de decisão . . . . .	30
4.2.2	Floresta aleatória . . . . .	31
4.2.3	Regressão logística . . . . .	32
4.2.4	<i>Support Vector Machine</i> . . . . .	33
4.2.5	Comparação dos resultados obtidos pelo algoritmo de melhor desempenho com o critério de risco utilizado pela Defesa Civil . . . . .	33
<b>5</b>	<b>CONSIDERAÇÕES FINAIS . . . . .</b>	<b>36</b>
	<b>Referências . . . . .</b>	<b>37</b>

# 1 Introdução

## 1.1 Justificativas e Relevância

A ocorrência de deslizamentos de terra tem sido responsável por uma série de problemas sociais ao longo dos anos, incluindo a destruição de edificações e infraestruturas urbanas, prejuízos às atividades produtivas, impactos ambientais e perda de vidas humanas (IBGE, 2019).

De acordo com um estudo realizado pela Companhia de Pesquisa de Recursos Minerais - CPRM no ano de 2021, foram mapeadas 313 áreas de risco geológico no município de Ouro Preto - MG, onde residem, aproximadamente, 3006 pessoas, representando cerca de 4% da população local. Destas áreas de riscos, mais de 80% apresentam suscetibilidade a perdas ou danos causados pela instabilização de encostas a partir da deflagração de deslizamentos (LANA; MARCUSSI, 2021).

Diante do exposto, torna-se indispensável lançar mão de mecanismos que possam atuar de forma a antecipar a ocorrência destes fenômenos, auxiliando o poder público na tomada de decisões que possam resguardar bens materiais e vidas.

Para tanto, este trabalho visa avaliar a capacidade de diferentes algoritmos de aprendizagem em máquina predizer eventos de movimento de massa no município de Ouro Preto - MG, tendo como base registros pretéritos destes eventos e os relacionando com o acumulado pluviométrico no mesmo período.

## 1.2 Objetivos

O principal objetivo do presente estudo é verificar a capacidade de predição de diferentes algoritmos de *machine learning* quanto ao risco de movimentações de massa no município de Ouro Preto - MG.

Derivados do Objetivo Geral apresentado, surgem os Objetivos Específicos que nortearam a presente pesquisa:

- Coletar e tratar os dados de movimentos de massa registrados pela Defesa Civil de Ouro Preto;
- Comparar a acurácia dos algoritmos desenvolvidos, utilizando no modelo final o algoritmo de melhor desempenho;

- Colacionar o critério atual de nível de alerta para risco de deslizamentos utilizado pela Defesa Civil com o resultado do modelo final.

## 2 Revisão de literatura

### 2.1 *Machine Learning*

O conceito de *machine learning*, em português aprendizado de máquina, pode ser definido como a área de estudo que tem como objetivo reconhecer padrões em conjuntos de dados submetidos à análise. Mais especificamente, ela trata do desenvolvimento de algoritmos que aprendem com os dados analisados e a partir destes são capazes de realizar previsões (NELLI, 2015).

Por vezes, interpretar os dados coletados torna-se uma atividade difícil. Para estes casos, o aprendizado de máquina é uma ferramenta imprescindível. Devido à grande quantidade de conjuntos de dados disponíveis, há um aumento crescente na demanda por essa tecnologia. Diversas indústrias estão empregando técnicas de aprendizado de máquina para extrair informações relevantes e vários estudos vêm sendo conduzidos para desenvolver métodos que permitam às máquinas aprenderem de forma autônoma, sem necessidade de programação explícita (MAHESH, 2020).

Há diversos métodos associados ao aprendizado de máquina, cada um apresentando características distintas que se alinham à natureza dos dados e ao modelo preditivo a ser construído. A decisão sobre qual método aplicar constitui o cerne do problema de aprendizagem. Os dados a serem processados durante a fase de aprendizagem podem se apresentar como matrizes contendo valores únicos por elemento ou valores multivariados (NELLI, 2015).

#### 2.1.1 Árvore de Decisão

As árvores de decisão são modelos sequenciais que utilizam uma série de testes simples e lógicos para combinar informações. Cada teste compara um atributo numérico com um valor limite ou um atributo nominal com um conjunto de valores possíveis. Esses classificadores simbólicos possuem a vantagem de serem muito mais compreensíveis em relação aos modelos de "caixa-preta", como as redes neurais (KOTSIANTIS, 2013).

Um exemplo de estrutura básica de uma árvore de decisão pode ser observado na figura 1. Esta estrutura, geralmente, se inicia a partir de um único nó que se ramifica em seus resultados possíveis. Essas ramificações repetem o comportamento do nó inicial e, como consequência, geram novas ramificações criando o aspecto de árvore que dá nome ao algoritmo.



Figura 1 – Árvore de decisão.

### 2.1.2 Floresta Aleatória

A floresta aleatória é um algoritmo de *machine learning* supervisionado. Por ser um algoritmo simples, confiável e flexível, a sua utilização é amplamente requerida. Devido a sua versatilidade em poder ser usado tanto para tarefas de classificação quanto para regressão, combinado com sua não linearidade, faz com que a floresta aleatória seja um algoritmo altamente adaptável a uma grande variedade de dados e situações (TIBCO, 2023).

O algoritmo floresta aleatória é formado pelo agrupamento de árvores de decisões, e cada árvore que a compõe contribui igualmente na predição final. A predição final deste modelo se dá de acordo com o objetivo do mesmo, classificação ou regressão. Na classificação esta predição é definida através do voto majoritário, enquanto na regressão, o resultado final é uma média entre todas as decisões individuais (PONTE; CAMINHA; FURTADO, 2020). A estrutura genérica da floresta aleatória e os cenários das predições finais deste modelo podem ser vistos na figura 2.

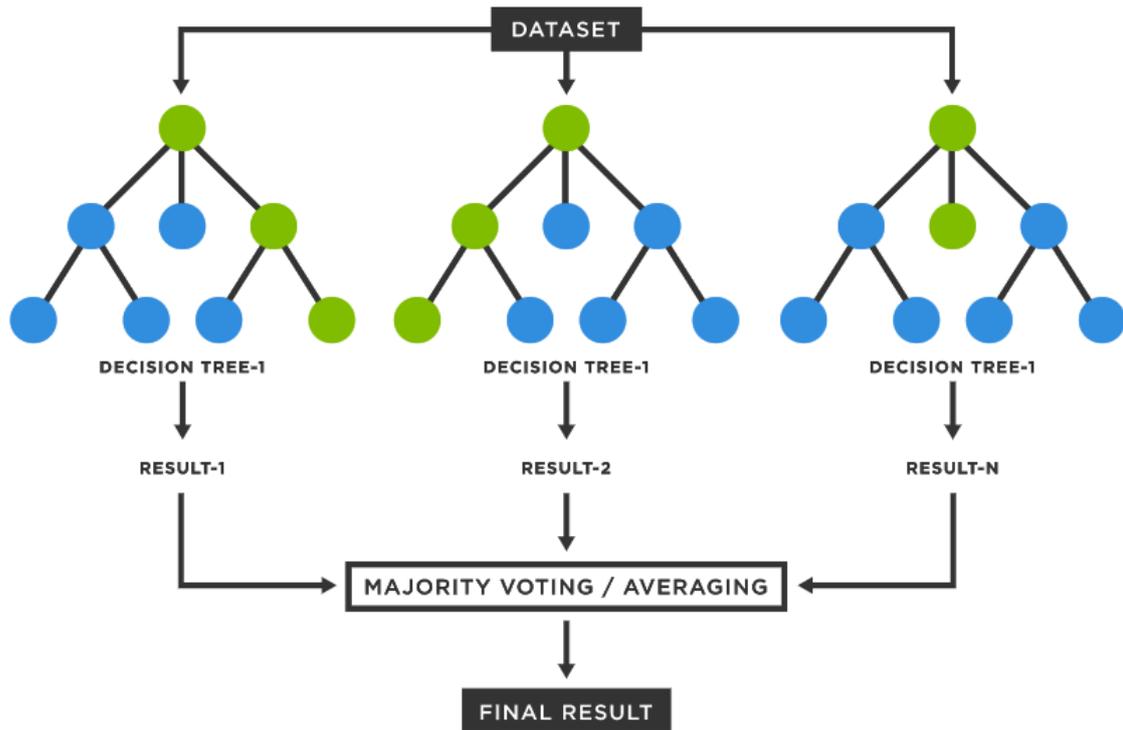


Figura 2 – Floresta aleatória. Fonte: (TIBCO, 2023)

### 2.1.3 Regressão Logística

Na regressão logística, a variável de resposta adota exclusivamente valores 0 e 1, onde '1' geralmente representa a ocorrência do evento de interesse, enquanto '0' indica a ausência desse evento (SILVEIRA et al., 2021). A relação entre as variáveis independentes e a variável dependente é denominada curva logística e seu comportamento se dá conforme a figura 3.

O modelo de regressão logística estima probabilidades para prever os resultados. Caso a probabilidade prevista seja maior do que 0,50, então a previsão é de que o resultado seja 1, isto é, a ocorrência do evento, e caso contrário, a previsão será 0, significando a não ocorrência (HAIR et al., 2009).

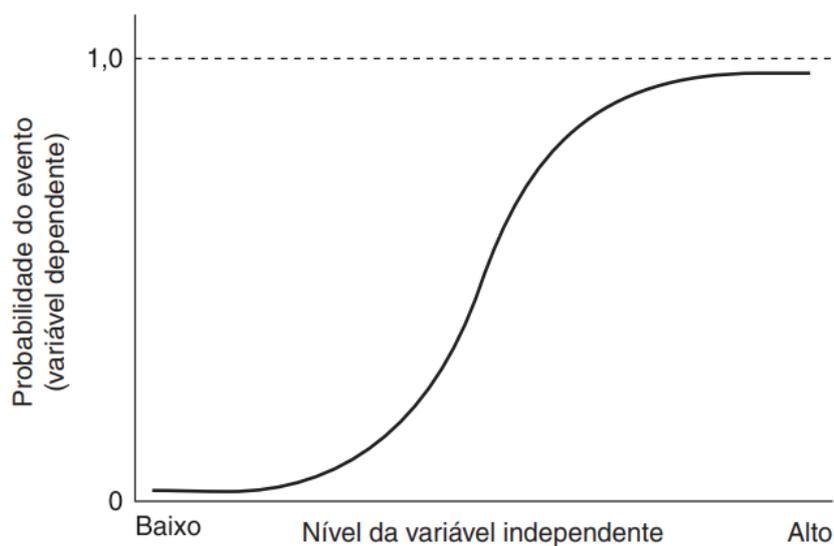


Figura 3 – Curva logística. Fonte: (HAIR et al., 2009)

#### 2.1.4 Support Vector Machine

Desenvolvida na década de 1990, o *Support Vector Machine* é um algoritmo de aprendizagem em máquina supervisionado que realiza a classificação dos dados traçando uma linha ou hiperplano ideal, conforme visto na figura 4, que maximiza a distância existente entre as diferentes classes (IBM, 2023). Após traçar esta reta, o algoritmo será capaz de prever a classe de novos dados de acordo com o lado em que estes forem alocados.

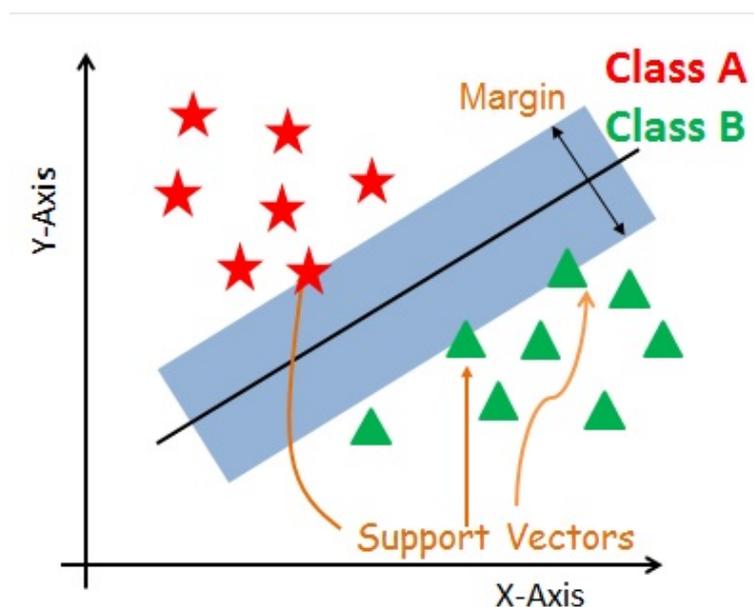


Figura 4 – Separação de classes no modelo SVM. Fonte: (DATACAMP, 2019)

O modelo SVM se desempenha melhor em modelos de classificação onde os dados

não estão distribuídos de maneira regular, uma vez que a separação dos dados pode ocorrer de forma não linear e independente. No entanto, a medida em que o conjunto de dados aumenta de tamanho, o tempo necessário para os cálculos executados pelo algoritmo também aumenta e o inverso ocorre com a interpretabilidade dos resultados, tornando-os de difícil compreensão (COUTINHO, 2019).

## 2.1.5 Métricas de desempenho

### 2.1.5.1 Matriz de confusão

A Matriz de Confusão é uma ferramenta utilizada no aprendizado de máquinas para avaliar a qualidade dos modelos, e para isso, através de uma matriz, demonstra a quantidade de acertos e erros de um algoritmo de classificação (AWARI, 2023).

Via de regra, a matriz de confusão é composta por quatro células:

- Verdadeiro positivo (VP): amostras que foram corretamente classificadas como pertencentes à classe positiva.
- Falso positivo (FP): amostras que foram erroneamente classificadas como pertencentes à classe positiva, quando na verdade pertenciam à classe negativa.
- Verdadeiro negativo (VN): amostras que foram corretamente classificadas como pertencentes à classe negativa.
- Falso negativo (FN): amostras que foram erroneamente classificadas como pertencentes à classe negativa, quando na verdade pertenciam à classe positiva.

Devido às suas características, a matriz de confusão se torna uma ferramenta valiosa para avaliar o desempenho de um modelo de classificação, permitindo uma compreensão clara de suas capacidades e limitações em relação aos diferentes tipos de classificações possíveis.

### 2.1.5.2 Acurácia

A acurácia é uma métrica que indica a performance geral do modelo, ou seja, dentre todas as predições, quantas o modelo classificou corretamente.

Com base na matriz de confusão, podemos obter a Acurácia como demonstrado na fórmula 2.1:

$$Acurácia = \frac{VP + VN}{VP + FP + VN + FN} \quad (2.1)$$

### 2.1.5.3 Precisão

A Precisão é a métrica que indica a proporção de verdadeiros positivos em relação ao total de dados classificados como positivo pelo modelo. A Precisão é inversamente proporcional ao número de classificações falso-positivas.

Com base na matriz de confusão, podemos obter a Precisão como demonstrado na fórmula 2.2:

$$Precisão = \frac{VP}{VP + FP} \quad (2.2)$$

### 2.1.5.4 Recall

O *Recall* é a métrica que mede a proporção de verdadeiros positivos em relação ao total de dados positivos contidos no modelo. É uma métrica de extrema importância quando a predição de falsos negativos é prejudicial (DUARTE, 2023).

Com base na matriz de confusão, podemos obter o *Recall* como demonstrado na fórmula 2.3:

$$Recall = \frac{VP}{VP + FN} \quad (2.3)$$

### 2.1.5.5 F1-score

A métrica *F1-score* pode ser definida como a média harmônica da precisão e do *recall*, com resultados que variam de 0 a 1, sendo o primeiro o pior cenário e o segundo o melhor (SKLEARN, 2023).

Podemos obter o *F1-score* como demonstrado na fórmula 2.4:

$$F1-score = 2 \times \left( \frac{Recall \times Precisão}{Recall + Precisão} \right) \quad (2.4)$$

## 2.2 Movimentos de massa

Os movimentos de massa, conhecidos também como deslizamento ou escorregamento, são movimentos de desprendimento de solos e rochas sob o efeito da gravidade, que geralmente tem seus efeitos agravados quando em combinação com eventos pluviais (CEMADEN, 2022).

Segundo o Cemaden (2022), os movimentos de massa são divididos em quatro grupos principais:

- **Quedas, tombamentos e rolamentos:**

As quedas são movimentos descendentes de fragmentos rochosos de volumes variados que se soltam, em queda livre, de taludes íngremes. Já o tombamento, é categorizado quando um bloco rochoso realiza um movimento de rotação frontal para fora do talude. Os rolamentos, por outro lado, são movimentos de blocos rochosos ao longo de encostas, frequentemente causados por descalçamentos.

- **Deslizamentos e escorregamentos:**

Os deslizamentos ou escorregamentos são movimentos de solo e rocha que ocorrem em superfícies de ruptura de talude. Estes movimentos podem ser classificados em deslizamento rotacional e translacional. O primeiro ocorre com materiais superficiais homogêneos e quando a superfície de ruptura é curvada no sentido superior com movimento rotatório. Já o segundo, ocorre em solos rasos e com origem em superfícies relativamente planas.

- **Fluxos de detritos e lama:**

São movimentos que têm como característica serem extremamente rápidos e com alto poder destrutivo, e que são desencadeados por um intenso fluxo de água na superfície que os originaram. Este intenso fluxo liquefaz o material superficial transformando-o em uma massa viscosa composta por lama e detritos rochosos, que escoam pela encosta.

- **Subsidência e colapsos:**

Estes movimentos ocorrem quando há o afundamento rápido ou gradual do solo, originado pelo colapso de cavidades, diminuição da porosidade do solo ou deformação de material argiloso.

As condições climáticas e geomorfológicas presentes no Brasil, como chuvas intensas de verão e presença de grandes maciços montanhosos, favorecem a ocorrência de movimentos de massa. Além dos fatores naturais, outros agentes contribuem para a ocorrência destes fenômenos e, seguramente, o mais prevalente destes fatores são as constantes ações antrópicas. A ocupação desordenada e irregular nas encostas é a ação antrópica que mais gera preocupação, uma vez que esta acarreta na supressão da cobertura vegetal e cortes de taludes, que por sua vez, aumentam a probabilidade de ocorrência dos movimentos de massa (GUIMARÃES et al., 2008).

## 2.3 Local de estudo

### 2.3.1 Localização geográfica

O município de Ouro Preto está localizado na região da Serra do Espinhaço no estado de Minas Gerais, distante cerca de 97km de Belo Horizonte, capital do estado. Sua

área territorial se estende por 1.245,865 km<sup>2</sup>, sendo ocupada por uma população estimada em 74.821 pessoas (IBGE, 2022).

### 2.3.2 Climatologia

O clima de Ouro Preto é caracterizado, segundo a classificação publicada por Köppen e Geiger<sup>1</sup>, como um clima subtropical de altitude (Cwb)<sup>2</sup>. Clima de comum ocorrência em regiões serranas brasileiras, encontradas, sobretudo, na região sudeste do país.

Fruto de sua condição altimétrica, o município de Ouro Preto apresenta médias de temperatura amenas que giram próximo dos 18,5<sup>o</sup> C, apresentando seu pico de temperatura durante o mês de janeiro e suas menores temperaturas são comumente registradas durante o mês de julho (CASTRO, 2006).

O nível médio de precipitação anual no município de Ouro Preto, é classificado como alto, apresentando os maiores índices pluviométricos nos meses de outubro a março, condição climática típica de regiões tropicais por apresentar verões significativamente quentes e demasiadamente chuvosos.

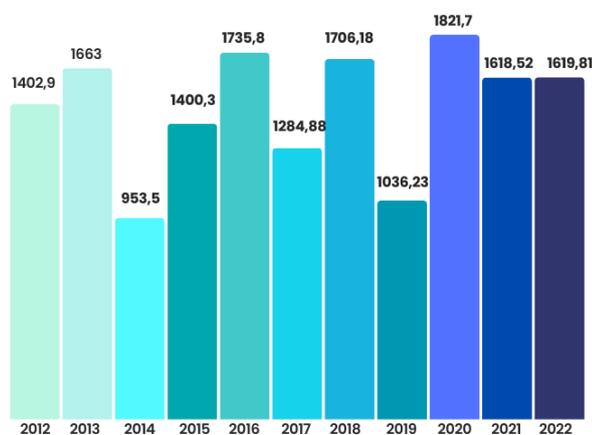


Figura 5 – Gráfico do acumulado anual pluviométrico no município de Ouro Preto nos anos de 2012 a 2022.

Os dados pluviométricos do município podem ser observados na figura 5, que apontam para uma média pluviométrica anual de 1476,62mm; média de precipitação que atualiza os dados publicados pela pesquisa de Castro (2006), que apontavam uma média pluviométrica de 1610,1mm para o período de 1988 a 2004.

<sup>1</sup> Classificação climática de referência para estudos na geografia e em suas subáreas desenvolvida por Wladimir Köppen (1990) e aperfeiçoada por Rudolf Geiger (1936).

<sup>2</sup> Sigla de referência internacional desenvolvida por Köppen-Geiger que representa *Temperate/Dry winter/Warm summer* (Temperado/Inverno seco/Verão quente).

### 2.3.3 Topografia e declividade

A serra em que se localiza o município está assentada/disposta na vertente sul da estrutura geológica denominada anticlinal de Mariana; caracterizada por possuir aspectos morfológicos onde se destacam a presença de vertentes íngremes e salientes e vales profundos e alongados (SOBREIRA; FONSECA, 2001).

Em toda a extensão do município de Ouro Preto é possível notar a presença de um relevo oscilante, apresentando em regiões como as das serras Ouro Preto e Itacolomi altitudes entre os 1200 e 1500 metros com grandes oscilações alcançando a marca de 1060 metros. A região do Ribeirão do Funil é o ponto de menor altitude no município, com valores inferiores à 1000 metros (FONTES, 2011).

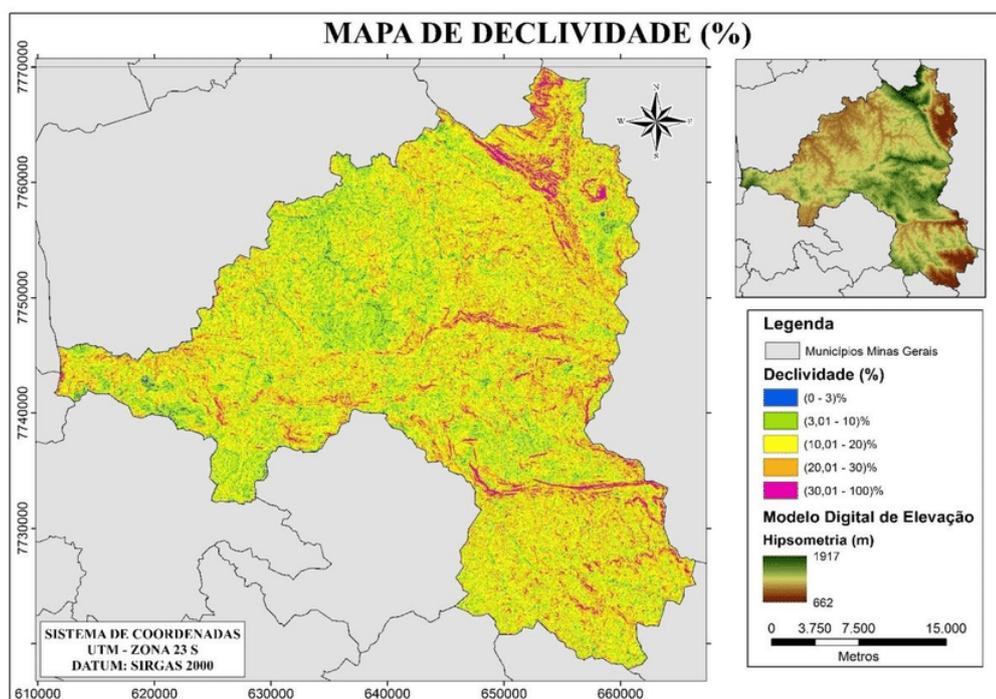


Figura 6 – Mapa de declividade e hipsometria do município de Ouro Preto Fonte: (PAMPOLINI; LANA, 2017)

### 2.3.4 Áreas de risco

Segundo o "Mapeamento de riscos em encostas e margem de rios" dos autores Carvalho, Macedo e Ogura (2007), o conceito de áreas de risco pode ser definido como:

Área passível de ser atingida por fenômenos ou processos naturais e/ou induzidos que causem efeitos adversos. As pessoas que habitam essas áreas estão sujeitas a danos à integridade física, perdas materiais e patrimoniais.

No mapa, visto na figura 7, podemos observar os pontos de susceptibilidade a escorregamentos no município de Ouro Preto. Destaca-se a expressiva predominância de zonas de alta e média susceptibilidade a este evento. Este cenário pode ser justificado pelo mapa de declividade demonstrado na figura 6, onde o relevo do município de Ouro Preto, apresenta, majoritariamente, declividades que variam de 10,01 a 20%, com zonas de alta declividade que variam de 30,01 a 100%. São estas zonas de alta declividade associadas a uma ocupação irregular e desordenada das encostas, que tornam o município propenso a eventos de movimentos de massa.

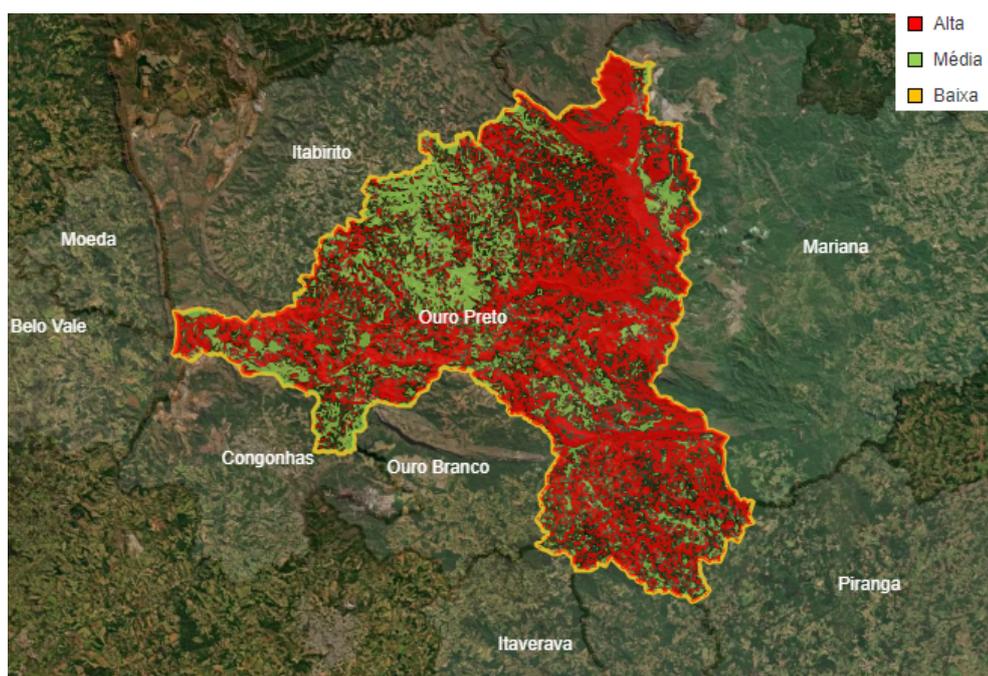


Figura 7 – Mapa de susceptibilidade a escorregamentos no município de Ouro Preto  
Fonte:([PROMOSAT-OP, 2024](#))

## 3 Materiais e Métodos

### 3.1 Coleta dos dados

O presente trabalho utilizou dados pluviométricos coletados pela Defesa Civil de Ouro Preto - MG no espaço temporal compreendido entre os anos de 2012 a 2022. Os dados de movimentações de massa foram retirados dos relatórios anuais da Defesa Civil, para este mesmo período, onde de um total de 1770 relatórios, 116 representavam registros do evento de interesse, 6,55% do total analisado.

### 3.2 Pré-processamento dos dados

Foram realizadas técnicas de pré-processamento para tratar as informações obtidas, de maneira a evitar vieses que afetassem a capacidade preditiva dos algoritmos. Desta forma, dezessete registros de interesse foram descartados por apresentarem inconsistências, como: não possuírem data ou por não estarem relacionados à períodos chuvosos. Como critério de exclusão, foram considerados como não relacionados aos períodos chuvosos dados de movimentos de massa que possuíam 0mm de acumulado pluviométrico em cinco dias.

Os registros removidos, conforme a figura 8, representam 14,65% do total de eventos de interesse. Já nos dados pluviométricos, foram verificadas as presenças de *outliers* e de dados faltantes, não havendo necessidade de correções.

Período	Mês	Precipitação/dia	Acumulado 5 dias	Dias consecutivos com precipitação	Ocorrência de movimento de massa
01/01/12	1	88,0	206,0	2	0
02/01/12	1	75,0	236,0	3	0
03/01/12	1	45,0	240,0	4	1
04/01/12	1	0,0	240,0	0	0
05/01/12	1	0,0	208,0	0	1
06/01/12	1	27,0	147,0	1	1
07/01/12	1	98,0	170,0	2	0
08/01/12	1	17,5	142,5	3	0
09/01/12	1	37,0	179,5	4	1
10/01/12	1	17,0	196,5	5	0
11/01/12	1	0,8	170,3	6	1
12/01/12	1	15,0	87,3	7	0
13/01/12	1	0,0	69,8	0	0
14/01/12	1	0,0	32,8	0	0
15/01/12	1	0,0	15,8	0	0
16/01/12	1	0,0	15,0	0	0
17/01/12	1	0,0	0,0	0	0
18/01/12	1	0,0	0,0	0	0
19/01/12	1	0,0	0,0	0	0

Figura 8 – Dado excluído durante a etapa de pré-processamento

Após a etapa de pré-processamento, os dados pluviométricos referentes aos dez anos analisados foram alinhados com os dados de movimentação de massa, para serem utilizados na implementação dos algoritmos.

### 3.3 Algoritmos

Para implementação dos algoritmos foi utilizada a linguagem de programação *Python* através da biblioteca *scikit-learn* na plataforma *Google Colaboratory*.

Como entrada, os algoritmos receberam os parâmetros: acumulado pluviométrico de 24 horas, acumulado pluviométrico de 120 horas, dias consecutivos com precipitação e período dos eventos. A saída dos algoritmos foi a ocorrência ou não de movimentos de massa.

O conjunto de dados foi dividido em 80% das amostras para treinamento e 20% para teste. No conjunto de treinamento foi realizada a validação cruzada utilizando o método *k-fold* com dez camadas.

Foram realizados cenários de treinamento dos algoritmos com os dados brutos e balanceados. Para aumentar a capacidade preditiva dos algoritmos utilizados, foi realizado o balanceamento dos dados pelo método de subamostragem através da função *RandomUnderSampler*.

### 3.4 Métricas de desempenho

Para avaliar a capacidade preditiva dos algoritmos testados foram utilizados os seguintes métodos:

- Matriz de confusão;
- *Recall*;
- Acurácia;
- *F1-score*;
- Precisão.

Sendo o *recall* da classe de interesse o parâmetro de maior peso na definição de desempenho do melhor algoritmo.

## 4 Resultados e Discussão

### 4.1 Dados brutos

Na figura 9 temos a distribuição dos registros dos eventos de movimentos de massa durante o período analisado. Podemos observar que o evento de interesse representa uma parcela mínima do total de registros (2,1%), o que era esperado, uma vez que se trata de uma ocorrência inusual.

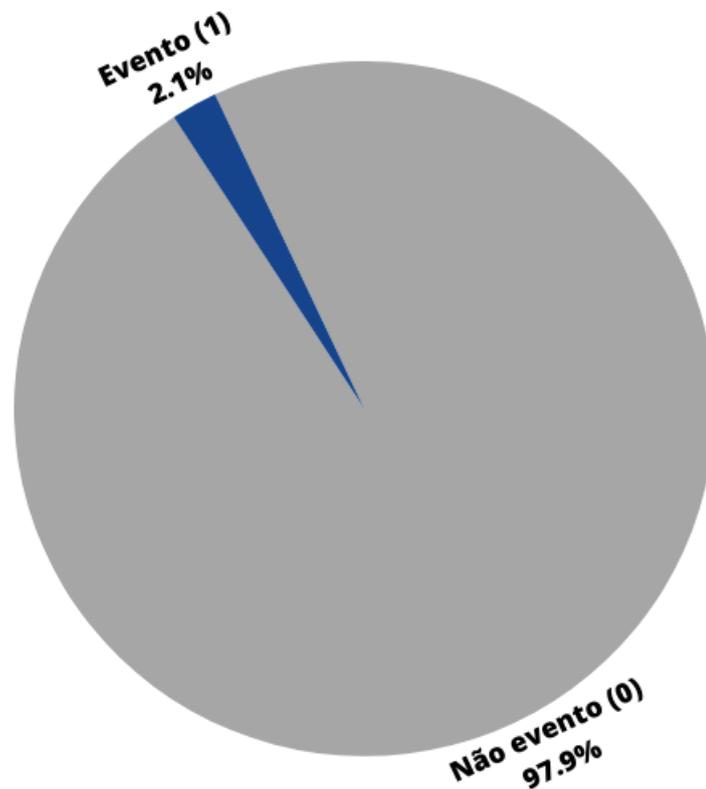


Figura 9 – Distribuição dos dados brutos.

#### 4.1.1 Árvore de decisão

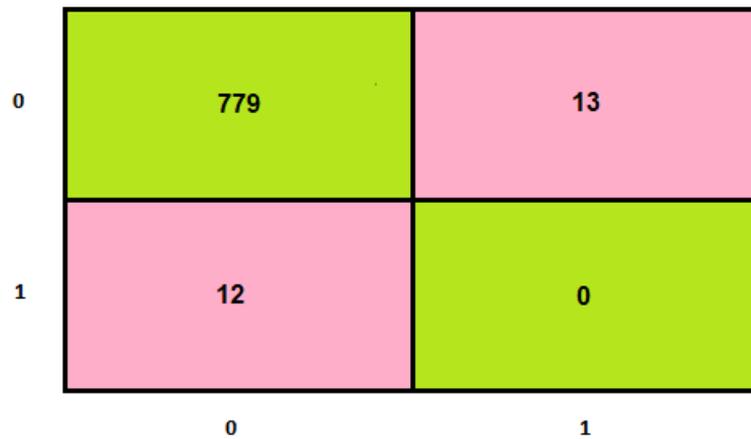


Figura 10 – Matriz de confusão do algoritmo árvore de decisão com dados brutos.

	Precisão	Recall	F1-score	Acurácia
0	0,98	0,98	0,98	0,97
1	0,00	0,00	0,00	
Média Ponderada	0,97	0,97	0,97	

Tabela 1 – Tabela das métricas de desempenho do algoritmo árvore de decisão com dados brutos.

Os resultados obtidos para a árvore de decisão são demonstrados na figura 10 e na tabela 1. O algoritmo, como demonstrado em sua matriz de confusão, teve 779 verdadeiros negativos, 12 falsos negativos, 13 falsos positivos e 0 verdadeiros positivos, resultando em uma precisão, *recall* e *F1-score* de 0,98 para a classe 'não-evento' e de 0,00 para a classe 'evento', com uma acurácia de 0,98.

#### 4.1.2 Floresta aleatória

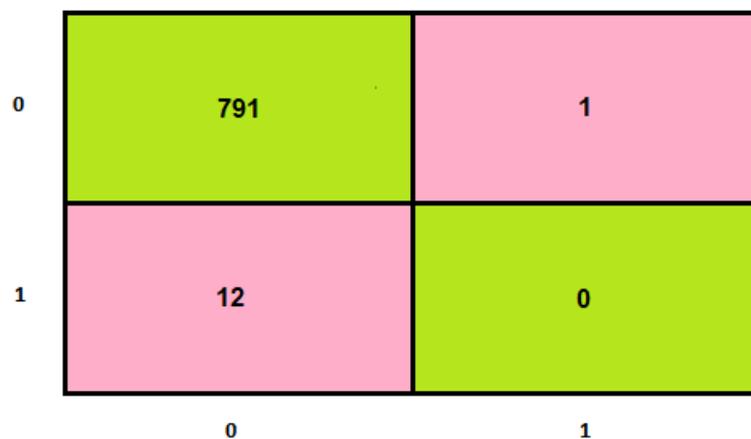


Figura 11 – Matriz de confusão do algoritmo floresta aleatória com dados brutos.

	Precisão	Recall	F1-score	Acurácia
0	0,99	1,00	0,99	0,98
1	0,00	0,00	0,00	
Média Ponderada	0,97	0,98	0,98	

Tabela 2 – Tabela das métricas de desempenho do algoritmo floresta aleatória com dados brutos.

Para o algoritmo floresta aleatória, vemos em sua matriz de confusão e nas métricas de desempenho, figura 11 e tabela 2, respectivamente, resultados similares aos encontrados na árvore de decisão com: 791 verdadeiros negativos, 12 falsos negativos, 1 falso positivo e 0 verdadeiros positivos, resultando em uma precisão de 0,99, *recall* de 1,00 e *F1-score* de 0,99, para a classe 'não-evento', e 0,00 para os três parâmetros na classe 'evento', com acurácia de 0,98.

### 4.1.3 Regressão logística

0	791	1
1	12	0
	0	1

Figura 12 – Matriz de confusão do algoritmo regressão logística com dados brutos.

	Precisão	Recall	F1-score	Acurácia
0	0,99	1,00	0,99	0,98
1	0,00	0,00	0,00	
Média Ponderada	0,97	0,98	0,98	

Tabela 3 – Tabela das métricas de desempenho do algoritmo de regressão logística com dados brutos.

Já para a regressão logística, temos novamente um cenário congênere aos dos modelos anteriormente apresentados, como visto na figura 12 e na tabela 3, com os seguintes resultados: 791 verdadeiros negativos, 12 falsos negativos, 1 falso positivo e 0 verdadeiros positivos, resultando em uma precisão de 0,99, *recall* de 1,00 e *F1-score* de

0,99, para a classe 'não-evento', e 0,00 para os três parâmetros da classe 'evento', com acurácia de 0,98.

#### 4.1.4 Support Vector Machine

0	792	0
1	12	0
	0	1

Figura 13 – Matriz de confusão do algoritmo SVM com dados brutos.

	Precisão	Recall	F1-score	Acurácia
0	0,99	1,00	0,99	0,99
1	0,00	0,00	0,00	
Média Ponderada	0,97	0,99	0,98	

Tabela 4 – Tabela das métricas de desempenho do algoritmo SVM com dados brutos.

Para o algoritmo SVM, os resultados novamente apresentam similaridades com os demais modelos, como demonstrado na figura 13 e na tabela 4, com: 792 verdadeiros negativos, 12 falsos negativos, 0 falsos positivos e 0 verdadeiros positivos, resultando em uma precisão de 0,99, *recall* de 1,00 e *F1-score* de 0,99, para a classe 'não-evento', e 0,00 para os três parâmetros para a classe 'evento', com acurácia de 0,99.

Este cenário de similaridade encontrado nos resultados dos algoritmos avaliados é justificado pelo fato da não realização do balanceamento dos dados, ou seja, os dados brutos apresentam uma grande discrepância entre as classes, sendo dominante a classe 'não-evento', como observado anteriormente na figura 9. Caso essa desproporcionalidade não seja considerada, os modelos podem sofrer do chamado "Paradoxo da Acurácia" onde a classe em menor proporção não consegue ser diferenciada pelos parâmetros dos algoritmos, uma vez que estes acreditam estar desempenhando um resultado satisfatório devido às altas acurácias alcançadas (AZANK, 2020), como observado nos resultados dos modelos apresentados, onde a acurácia mínima alcançada foi de 97% e, no entanto, nenhuma predição verdadeira-positiva foi atribuída pelos algoritmos.

## 4.2 Dados balanceados

O balanceamento dos dados, visto na figura 14, é uma técnica utilizada para nivelar o número de dados de interesse aos não, ou seja, deixar o ambiente de treinamento dos algoritmos, no caso do presente trabalho, com 50% de cada classe, facilitando o aprendizado dos mesmos.

Existem duas maneiras distintas de realizar a técnica de balanceamento, a primeira é a *undersampling*, que consiste em preservar os dados da classe com menor incidência e diminuir, de forma aleatória ou não, aqueles de maior frequência, equilibrando a variável alvo nas observações do conjunto de dados. Já a segunda, a *oversampling*, realiza a operação contrária, aumentando a classe de menor frequência até que ela atinja um nível de equilíbrio entre as classes do banco de dados (MIOTO; MIRANDA; PREMEBIDA, 2022).

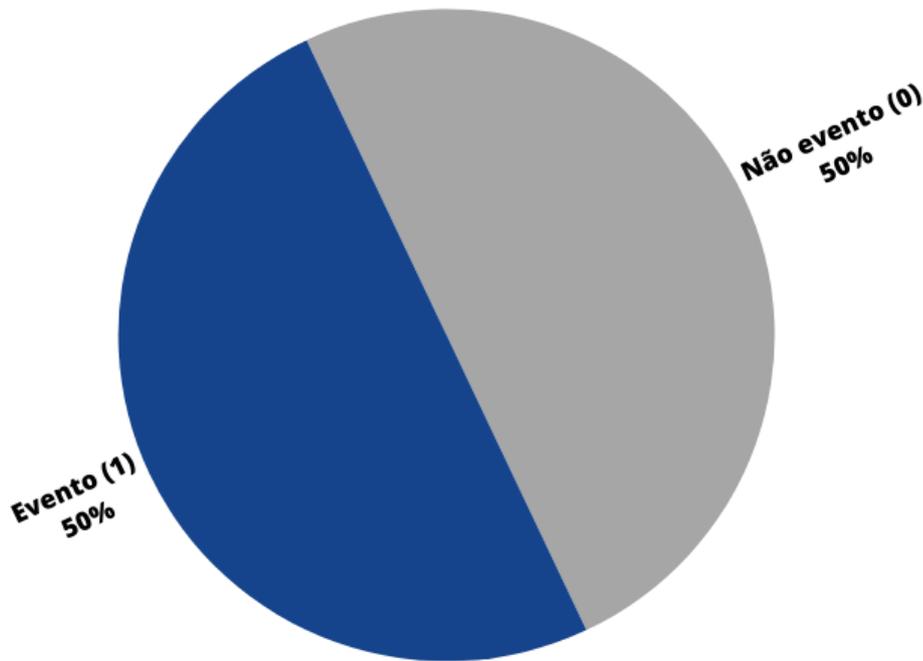


Figura 14 – Distribuição dos dados balanceados.

Para realizar o balanceamento dos dados deste trabalho, foi utilizado o método *undersampling* por este apresentar um desempenho melhor do que o *oversampling*, resultado observado também por Drummond et al. (2003) em seu trabalho que teve como objetivo comparar o desempenho destes dois métodos.

### 4.2.1 Árvore de decisão

0	560	232
1	2	10
	0	1

Figura 15 – Matriz de confusão do algoritmo árvore de decisão com balanceamento dos dados.

	Precisão	Recall	F1-score	Acurácia
0	1,00	0,71	0,83	0,71
1	0,04	0,83	0,08	
Média Ponderada	0,98	0,71	0,82	

Tabela 5 – Tabela das métricas de desempenho do algoritmo de árvore de decisão com balanceamento dos dados.

Para o algoritmo árvore de decisão, temos na figura 15 e na tabela 5, os seguintes resultados: 560 verdadeiros negativos, 2 falsos negativos, 232 falsos positivos e 10 verdadeiros positivos, resultando para classe 'não-evento' em uma precisão de 1,00, *recall* de 0,71 e *F1-score* de 0,83, e para a classe 'evento' os resultados foram de 0,04, 0,83 e 0,08, respectivamente para os mesmos parâmetros, com acurácia de 0,71.

#### 4.2.2 Floresta aleatória

0	593	199
1	2	10
	0	1

Figura 16 – Matriz de confusão do algoritmo floresta aleatória com balanceamento dos dados.

	Precisão	Recall	F1-score	Acurácia
0	1,00	0,73	0,85	0,74
1	0,05	0,83	0,09	
Média Ponderada	0,98	0,71	0,82	

Tabela 6 – Tabela das métricas de desempenho do algoritmo floresta aleatória com balanceamento dos dados.

Já o algoritmo árvore de decisão, apresentou os resultados demonstrados na figura 16 e na tabela 6, onde: 593 verdadeiros negativos, 2 falsos negativos, 199 falsos positivos e 10 verdadeiros positivos, resultando para classe 'não-evento' em uma precisão de 1,00, *recall* de 0,73 e *F1-score* de 0,85, e para a classe 'evento' os resultados foram de 0,05, 0,83 e 0,09, respectivamente para os mesmos parâmetros e acurácia de 0,74.

### 4.2.3 Regressão logística

0	666	126
1	5	7
	0	1

Figura 17 – Matriz de confusão do algoritmo de regressão logística com balanceamento dos dados.

	Precisão	Recall	F1-score	Acurácia
0	0,99	0,84	0,91	0,84
1	0,05	0,58	0,10	
Média Ponderada	0,98	0,84	0,90	

Tabela 7 – Tabela das métricas de desempenho do algoritmo de regressão logística com balanceamento dos dados.

Os resultados para o modelo de regressão logística, visto na figura 17 e na tabela 7, foram: 666 verdadeiros negativos, 5 falsos negativos, 126 falsos positivos e 7 verdadeiros positivos, resultando em uma precisão de 0,99, *recall* de 0,84 e *F1-score* de 0,91, para a classe 'não-evento', e 0,05, 0,58 e 0,10, respectivamente para os mesmos parâmetros para a classe 'evento', com acurácia de 0,84.

#### 4.2.4 Support Vector Machine

0	721	71
1	8	4
	0	1

Figura 18 – Matriz de confusão do algoritmo SVM com balanceamento dos dados.

	Precisão	Recall	F1-score	Acurácia
0	0,99	0,91	0,95	0,90
1	0,05	0,33	0,09	
Média Ponderada	0,98	0,90	0,94	

Tabela 8 – Tabela das métricas de desempenho do algoritmo SVM com balanceamento dos dados.

Por fim, temos, na figura 18 e na tabela 8, os resultados obtidos pelo algoritmo SVM, sendo: 721 verdadeiros negativos, 8 falsos negativos, 71 falsos positivos e 4 verdadeiros positivos, resultando em uma precisão de 0,99, *recall* de 0,91 e *F1-score* de 0,95, para a classe 'não-evento', e para a classe 'evento' os resultados foram 0,05, 0,33 e 0,09, respectivamente, para os mesmos parâmetros, com acurácia de 0,90.

Diferente do ocorrido para os dados brutos, nos dados balanceados, os algoritmos atingiram resultados e desempenhos particulares, com diferentes sensibilidades e especificidades. O algoritmo que melhor soube balancear estes dois parâmetros foi a floresta aleatória.

#### 4.2.5 Comparação dos resultados obtidos pelo algoritmo de melhor desempenho com o critério de risco utilizado pela Defesa Civil

A Defesa Civil adota um nível de alerta meteorológico que é atualizado de acordo com o acumulado pluviométrico registrado em um período de cinco dias. É considerado estado de alerta quando o acumulado atinge um valor de 128mm em um período de cinco dias consecutivos. Neste nível de alerta, movimentos de massa podem ocorrer isoladamente. No entanto, ao se analisar as características dos fenômenos ocorridos a partir dos dados coletados, foi observado que 70,93% dos movimentos de massa registrados ocorreram em níveis inferiores a 128mm de acumulado pluviométrico em cinco dias.

Já o algoritmo floresta aleatória apresentou um *recall* de 83%, ou seja, ela foi capaz de identificar 83% dos dados de movimentos de massa na fase de teste do modelo. Contudo, o algoritmo apresentou um valor elevado de falsos-positivos, que são dados equivocadamente interpretados como positivos pelo modelo, resultando em uma diminuição de sua acurácia.

	Média	Desvio padrão
<b>Precipitação diária</b>	13,72	15,08
<b>Acumulado de 5 dias</b>	56,60	35,07
<b>Dias consecutivos com precipitação</b>	3,90	4,23

Tabela 9 – Tabela de análise dos dados falsos-positivos do algoritmo floresta aleatória.

Por possuírem características de similaridade com os valores registrados nas ocorrências dos eventos, seria plausível imaginar uma correlação entre os dados falsos-positivos encontrados pelo modelo e o risco de movimentos de massa. No entanto, ao analisarmos a tabela 9, onde temos as médias e desvios padrões dos parâmetros pluviométricos para as avaliações falso-positivas fornecidas pelo algoritmo floresta aleatória, podemos observar uma alta dispersão dos dados, demonstrada pelos elevados valores de desvios padrões. Devido a esta dispersão, a correlação de risco de ocorrência de evento não é possível ser estabelecida, uma vez que os valores médios de precipitação diária e acumulado de cinco dias encontrados neste intervalo são recorrentes nos períodos chuvosos do município.

Estabelecer uma relação comparativa entre a metodologia de alerta utilizada pela Defesa Civil e os resultados obtidos pelo algoritmo floresta aleatória não é uma tarefa simples, uma vez que a primeira não tem como objetivo prever eventos de movimentos de massa, e sim estabelecer um cenário de risco para que providências cabíveis possam ser adotadas, enquanto o segundo busca entender os padrões existentes nos eventos para prevêê-los.

Por outro lado, é possível traçar um paralelo entre as circunstâncias que levaram os dois métodos a não obterem resultados satisfatórios em seus respectivos objetivos fins. O primeiro ponto diz respeito aos dados, uma vez que estes possuam possíveis inconsistências em sua origem, seja na etapa de coleta ou registro, a avaliação e o desempenho de qualquer metodologia que os envolvam terão seus resultados comprometidos. Na ciência de dados, temos uma máxima que diz: “*Garbage in – garbage out*”, em português, “entra lixo – sai lixo”, isto significa dizer que se a entrada de dados for de baixa qualidade, a saída resultante não será confiável, ou seja, as informações coletadas precisam ser precisas para que tenhamos uma saída fidedigna ao que foi proposto (KILKENNY; ROBINSON, 2018). O segundo ponto se refere à capacidade de, isoladamente, os parâmetros pluviométricos serem um indicador ideal para prever o risco de ocorrência de movimentos de massa, sendo necessário associar

este parâmetro a outros que aumentem a capacidade de generalização dos algoritmos e a compreensão dos mecanismos que envolvem a ocorrência destes movimentos no município.

## 5 Considerações Finais

A partir dos resultados obtidos no presente trabalho, é possível concluir que, apesar dos algoritmos não apresentarem resultados satisfatórios, suas utilizações, desde que alimentados com bases de dados sólidas, podem ser fator diferencial no auxílio dos órgãos públicos a tomarem decisões assertivas, como por exemplo, gerar uma nova classificação de alerta para o município. Ademais, se faz necessário munir tais órgãos com novas ferramentas de monitoramento que forneçam parâmetros além dos pluviométricos já utilizados, gerando dados que possam servir como base para futuros estudos, inclusive os que abarcam o mesmo tema que o presente, sendo possível fornecer aos algoritmos novas entradas e dando a eles a possibilidade de encontrarem novos padrões que justifiquem melhor a ocorrência de movimentos de massa no município de Ouro Preto - MG.

# Referências

AWARI. *Matriz de Confusão no TensorFlow: Avaliando a Precisão dos Modelos de Aprendizado de Máquina*. Awari. Jul. 2023. Disponível em: [https://awari.com.br/matriz-de-confusao-no-tensorflow-avaliando-a-precisao-dos-modelos-de-aprendizado-de-maquina/?utm\\_source=blog&utm\\_campaign=projeto+blog&utm\\_medium=Matriz%20de%20Confus%C3%A3o%20no%20TensorFlow:%20Avaliando%20a%20Precis%C3%A3o%20dos%20Modelos%20de%20Aprendizado%20de%20M%C3%A1quina](https://awari.com.br/matriz-de-confusao-no-tensorflow-avaliando-a-precisao-dos-modelos-de-aprendizado-de-maquina/?utm_source=blog&utm_campaign=projeto+blog&utm_medium=Matriz%20de%20Confus%C3%A3o%20no%20TensorFlow:%20Avaliando%20a%20Precis%C3%A3o%20dos%20Modelos%20de%20Aprendizado%20de%20M%C3%A1quina). Citado 1 vez na página 18.

AZANK, Felipe. *Dados Desbalanceados - O que são e como lidar com eles*. Medium. Set. 2020. Disponível em: <https://medium.com/turing-talks/dados-desbalanceados-o-que-s%C3%A3o-e-como-evit%C3%A1-los-43df4f49732b>. Citado 1 vez na página 29.

CARVALHO, Celso Santos; MACEDO, Eduardo Soares de; OGURA, Agostinho Tadashi. Mapeamento de riscos em encostas e margem de rios. *Brasília: Ministério das Cidades*, p. 2007, 2007. Citado 1 vez na página 22.

CASTRO, Jeanne Michelle Garcia. Pluviosidade e movimentos de massa nas encostas de Ouro Preto. Programa de Pós Graduação em Engenharia Civil. Departamento de Engenharia ..., 2006. Citado 2 vezes na página 21.

CEMADEN. *Movimento de Massa*. Centro Nacional de Monitoramento e Alertas de Desastres Naturais. Mar. 2022. Disponível em: <https://www.gov.br/cemaden/pt-br/paginas/ameacas-naturais/movimento-de-massa#:~:text=Movimento%20de%20Massa%2C%20tamb%C3%A9m%20denominado,potencializado%20pela%20a%C3%A7%C3%A3o%20da%20%C3%A1gua..> Citado 2 vezes na página 19.

COUTINHO, Bernardo. *Modelos de Predição / SVM*. Medium. Jul. 2019. Disponível em: <https://medium.com/turing-talks/turing-talks-12-classifica%C3%A7%C3%A3o-por-svm-f4598094a3f1>. Citado 1 vez na página 18.

DATA CAMP. *Support Vector Machines with Scikit-learn Tutorial*. Datacamp. 2019. Disponível em: <https://www.datacamp.com/tutorial/svm-classification-scikit-learn-python>. Citado 0 vez na página 17.

DRUMMOND, Chris et al. C4. 5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling. In: WORKSHOP on learning from imbalanced datasets II. 2003. v. 11, p. 1-8. Citado 1 vez na página 30.

DUARTE, Rafael. *Métricas de Avaliação em Modelos de Classificação em Machine Learning*. Sigmoidal. Nov. 2023. Disponível em: <https://sigmoidal.ai/metricas-de-avaliacao-em-modelos-de-classificacao-em-machine-learning/>. Citado 1 vez na página 19.

- FONTES, Michel Moreira Morandini. Contribuição para o desenvolvimento da metodologia de análise, gestão e controle de riscos geotécnicos para a área urbana da cidade de Ouro Preto. Programa de Pós-Graduação em Geotecnia. Núcleo de Geotecnia, Escola de Minas ..., 2011. Citado 1 vez na página 22.
- GUIMARÃES, Renato Fontes et al. Movimentos de massa. In: *Geomorfologia: conceitos e tecnologias atuais*. Edição: Teresa Gallotti Florenzano. Brasil: Oficina de textos, 2008. P. 320. Citado 1 vez na página 20.
- HAIR, Joseph F et al. *Análise multivariada de dados*. Bookman editora, 2009. Citado 1 vez nas páginas 16, 17.
- IBGE. *Cidades e Estados*. Instituto Brasileiro de Geografia e Estatística. 2022. Disponível em: <https://www.ibge.gov.br/cidades-e-estados/mg/ouro-preto.html>. Citado 1 vez na página 21.
- IBGE. *Suscetibilidade a deslizamentos do Brasil: primeira aproximação*. Coordenação de Recursos Naturais e Estudos Ambientais, 2019. Citado 1 vez na página 12.
- IBM. *What are support vector machines (SVMs)?* IBM. Dez. 2023. Disponível em: <https://www.ibm.com/topics/support-vector-machine>. Citado 1 vez na página 17.
- KILKENNY, Monique F; ROBINSON, Kerin M. *Data quality: "Garbage in-garbage out"*. v. 47. SAGE Publications Sage UK: London, England, 2018. P. 103–105. Citado 1 vez na página 34.
- KOTSIANTIS, S B. Decision trees: a recent overview. en. *Artif. Intell. Rev.*, Springer Science e Business Media LLC, v. 39, n. 4, p. 261–283, abr. 2013. Citado 1 vez na página 14.
- LANA, Julio Cesar; MARCUSSI, Maria Carolina Rodrigues. *Diagnóstico da população em áreas de risco geológico, Ouro Preto, MG*. Serviço Geológico do Brasil – CPRM, 2021. Citado 1 vez na página 12.
- MAHESH, Batta. Machine learning algorithms-a review. *International Journal of Science and Research (IJSR)*. [Internet], v. 9, n. 1, p. 381–386, 2020. Citado 1 vez na página 14.
- MIOTO, Ana Clara; MIRANDA, João Vitor de; PREMEBIDA, Sthefanie Monica. *Lidando com o desbalanceamento de dados*. Alura. Abr. 2022. Disponível em: <https://www.alura.com.br/artigos/lidando-com-desbalanceamento-dados#:~:text=undersampling%20e%20oversampling.-,Undersampling,com%20a%20vari%C3%A1vel%20alvo%20equilibrada..> Citado 1 vez na página 30.
- NELLI, Fabio. *Python data analytics: Data analysis and science using PANDAs, Matplotlib and the Python Programming Language*. Apress, 2015. Citado 2 vezes na página 14.
- PAMPOLINI, Diógenes Guilherme; LANA, Cláudio Eduardo. Uso de geoprocessamento para indicação de áreas favoráveis à construção de aterro sanitário no município de Ouro Preto (MG). *Caderno de Geografia*, v. 27, p. 641–655, 2017. Citado 0 vez na página 22.

PONTE, Caio; CAMINHA, Carlos; FURTADO, Vasco. Optimizing Random Forest from the pondering of regression tree leaves, 2020. Citado 1 vez na página 15.

PROMOSAT-OP. *WebMAPA do Programa de Monitoramento Socioambiental do Território Ouro-pretano*. Programa de Monitoramento Socioambiental do Território Ouro-pretano. Jan. 2024. Disponível em: <https://promosatop.ufop.br/>. Citado 0 vez na página 23.

SILVEIRA, Maria Beatriz Galdino da et al. Aplicação da regressão logística na análise dos dados dos fatores de risco associados à hipertensão arterial. *Research, Society and Development*, v. 10, n. 16, e20101622964–e20101622964, 2021. Citado 1 vez na página 16.

SKLEARN. *sklearn.metrics.f1\_score*. Sklearn. 2023. Disponível em: [https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1\\_score.html](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html). Citado 1 vez na página 19.

SOBREIRA, Frederico Garcia; FONSECA, Marco Antônio. Impactos físicos e sociais de antigas atividades de mineração em Ouro Preto, Brasil., 2001. Citado 1 vez na página 22.

TIBCO. *What is a random forest*. Tibco. 2023. Disponível em: <https://www.tibco.com/pt-br/reference-center/what-is-a-random-forest>. Citado 1 vez nas páginas 15, 16.