

COMPARANDO ALGORITMOS BASEADOS EM ÁRVORE PARA CLASSIFICAÇÃO DO RISCO DE FOGO EM ALTAMIRA-PA

COMPARING TREE-BASED ALGORITHMS FOR FIRE RISK CLASSIFICATION IN ALTAMIRA

Averaldo Juvenal Menezes Filho¹

Juan Carlos Santos Parente²

Laiane Martins de Souza³

Nicole Victória Lopes da Costa⁴

Marta de Oliveira Barreiros⁵

Área Temática: Meio ambiente, Mudanças Climáticas e Sustentabilidade

Modalidade: Artigo Científico

Resumo

A Amazônia tem enfrentado um número alarmante de focos de incêndio, conforme registros do Instituto Nacional de Pesquisas Espaciais (INPE), o que sublinha a urgência de uma abordagem mais precisa para essa problemática. Para isso, é fundamental compreender as variáveis que influenciam a ocorrência desses focos. Este estudo foca na comparação da eficiência de dois algoritmos de aprendizado de máquina baseados em sistemas de árvore – Random Forest e Extremely Randomized Trees (Extra Trees) na classificação do risco de fogo no município de Altamira, Pará. A escolha de Altamira justifica-se por ser um dos locais com alta incidência de incêndios registrados pelo INPE no estado. Para a análise, além dos dados históricos de incêndio, foram utilizados dados meteorológicos fornecidos pelo Instituto Nacional de Meteorologia (INMET). Ambos os algoritmos demonstraram eficiência notável na classificação do risco. O Random Forest alcançou uma acurácia de 90,4%, enquanto o Extra Trees

¹ Universidade do Estado do Pará; averaldo.jm.filho@aluno.uepa.br

² Universidade do Estado do Pará; juan.csparente@aluno.uepa.br

³ Universidade do Estado do Pará; laiane.md.souza@aluno.uepa.br

⁴ Universidade do Estado do Pará; nicole.costa@aluno.uepa.br

⁵ Universidade do Estado do Pará; marta.do.barreiros@uepa.br

obteve um desempenho ligeiramente superior, com 90,8% de acurácia. Esses resultados apontam para o potencial desses modelos na previsão e gestão do risco de incêndios na região.

Palavras-Chave: Altamira, Risco de Fogo, Aprendizado de Máquina, Pará, Amazônia

Abstract

The Amazon has been facing an alarming number of fire outbreaks, according to records from the National Institute for Space Research (INPE), which highlights the urgency of a more accurate approach to this problem. To this end, it is essential to understand the variables that influence the occurrence of these outbreaks. This study focuses on comparing the efficiency of two machine learning algorithms based on tree systems – Random Forest and Extremely Randomized Trees (Extra Trees) in classifying fire risk in the municipality of Altamira, Pará. Altamira was chosen because it is one of the locations with the highest incidence of fires recorded by INPE in the state. For the analysis, in addition to historical fire data, meteorological data provided by the National Institute of Meteorology (INMET) were used. Both algorithms demonstrated remarkable efficiency in classifying risk. Random Forest achieved an accuracy of 90.4%, while Extra Trees achieved a slightly higher performance, with 90.8% accuracy. These results point to the potential of these models in predicting and managing fire risk in the region.

Key words: Altamira, Fire Risk, Machine Learning, Pará, Amazon

1. Introdução

A Amazônia, segundo o Instituto Nacional de Pesquisas Espaciais (INPE), é o bioma brasileiro com o maior número de registros de focos de incêndio. Somente no estado do Pará, foram detectados 1.466 focos em 2024 (INPE, 2025). Os perigos desses incêndios impactam diretamente as populações humanas e o ecossistema. A fumaça, por exemplo, polui o ar, enquanto os resíduos do fogo contaminam o solo e a água (MARTIN et al., 2016). Essas consequências graves evidenciam a urgência de desenvolver métodos eficazes de previsão de risco para implementar soluções mais precisas.

O município de Altamira, no Pará, tem sido um dos locais com alta incidência de incêndios registrados pelo INPE. Diante desse cenário, este estudo propõe classificar o risco de ocorrência de incêndios nesse município em níveis baixo, médio e alto. Para isso, serão utilizados os algoritmos de aprendizado de máquina Random Forest e Extremely Randomized Trees (Extra Trees), que correlacionam os registros de incêndios com dados meteorológicos locais, fornecidos pelo Instituto Nacional de Meteorologia (INMET).

Segundo dados do Monitor de Fogo do MapBiomas, mais de 30,8 milhões de hectares foram queimados no Brasil em 2024, o que representa um aumento de 79% em relação ao ano de 2023. Desse total, mais da metade corresponde ao bioma amazônico.

Índices de risco de fogo têm sido utilizados como ferramentas preditivas para identificar áreas vulneráveis. No estudo de Galvão Junior et al. (2022), principal referência deste trabalho, esses índices combinam variáveis como dias sem chuva, ocorrência de precipitação e registros de incêndios, permitindo a classificação do risco em níveis de alerta por meio de sistemas neuro-fuzzy.

Tan, C. e Feng, Z. (2023) utilizaram informações meteorológicas, topográficas, de vegetação e socioeconômicas para criar mapas de risco de fogo na província de Hunan, na China. Ao comparar os algoritmos de aprendizado de máquina Random Forest (RF), Support Vector Machine (SVM) e Gradient Boosting Decision Tree (GBDT), o estudo concluiu que o RF foi o mais eficaz, alcançando uma precisão de 91,68% com os dados analisados.

Chang; Yoon e Lee (2020) compararam a eficiência de modelos de Rede Neural, Decision Tree, XGBoost e Extra Trees para classificar riscos de incêndios na província de Gyeongnam, na Coreia do Sul, resultando num destaque do Random Forest performando 89.3% de acurácia.

Na Grécia, Maniatis, Y.; Doganis, A.; e Chatzigeorgiadis, M. (2022) buscaram desenvolver um modelo de risco de incêndio para o Parque Florestal Nacional de Dadia-Lefkimi-Soufli, utilizando ferramentas de aprendizado de máquina e análise de decisão multicritério em um ambiente de Sistema de Informação Geográfica (SIG). Com base em fatores como cobertura do solo, altitude, declividade e índice de umidade topográfica, o objetivo principal foi mapear áreas de alto risco de incêndio e analisar o impacto das mudanças na cobertura vegetal. Como resultado, demonstraram a utilidade do modelo na gestão de incêndios e na alocação de recursos.

2. Metodologia

Para a metodologia do presente trabalho, adotou-se um índice de classificação de risco de fogo desenvolvido com base na cidade de Sorocaba, em São Paulo, conforme proposto por Galvão Junior et al. (2022). No estudo de Sorocaba, o autor elabora o referido índice e implementa um algoritmo baseado em lógica neuro-fuzzy, empregando dados oficiais do Instituto Nacional de Pesquisas Espaciais (INPE). Com os dados fornecidos pelo instituto, é possível compreender a radiação térmica das localidades, o que permite a classificação dos riscos de incêndio na região de estudo (JUNIOR, 2022, p. 3). Para este trabalho, foram utilizados os dados sobre focos de incêndio no Brasil fornecidos pelo INPE, além de dados sobre temperatura disponibilizados pelo Instituto Nacional de Meteorologia (INMET).

Diante do grande volume de informações disponíveis, foi necessário aplicar a filtragem de dados para manter apenas o que seria relevante ao estudo. O município de Altamira foi o escolhido para a realização do trabalho, utilizando dados obtidos pelo satélite GOES-16 nos anos de 2022, 2023 e 2024. O treinamento do software foi realizado utilizando a linguagem Python na versão 3.10.12, e os algoritmos escolhidos para a criação dos modelos de classificação de dados foram os algoritmos de aprendizado de máquina Random Forest e Extra Trees, por já serem ferramentas utilizadas e conferidas na literatura consultada para o desenvolvimento desta publicação.

2.1. Base de Dados

Os dados do estudo foram obtidos por meio do portal “TerraBrasilis” (<https://terrabrasilis.dpi.inpe.br/>) e do Banco de Dados Meteorologia do INMET (<https://bdmep.inmet.gov.br/>) que fornecem, respectivamente, os dados de queimadas do INPE e os dados de temperatura de diversas regiões do Brasil. A base de dados de registros de focos de incêndios florestais do INPE possuem as seguintes variáveis: DataHora, Satelite, País, Estado, Município, Bioma, DiaSemChuva, Precipitação, RiscoFogo, Latitude, Longitude e FRP(Fire Radiative Power ou Potência Radiativa do Fogo). As variáveis do INMET sobre temperatura são: Data,Hora, Temp. Ins., Temp. Max., Temp. Min., Umi. Ins., Umi. Max., Umi. Min., Pto Orvalho Ins., Pto Orvalho Max., Pto Orvalho Min., Pressao Ins., Pressao Max., Pressao Min., Vel. Vento, Dir. Vento, Raj. Vento, Radiação, Chuva e Cidade. A soma de todas essas variáveis resultou em um conjunto com um pouco mais de 21.000 dados.

2.2. Limpeza e Processamento de Dados

Para possibilitar o treinamento com dados consistentes foram selecionadas as seguintes variáveis do INMET: Data, Hora (UTC), Temperatura Máxima, Temperatura Mínima, Umidade Máxima, Umidade Mínima, Pressão Máxima, Pressão Mínima e Velocidade do Vento. Já do INPE, foram utilizadas as variáveis Data e Hora, Risco de Fogo, FRP, Dias sem Chuva e Precipitação.

Em seguida, foram eliminados os valores nulos dessas bases, para então proceder com a separação da data e da hora contidas na variável DataHora da base do INPE. A hora foi convertida para o horário UTC (Tempo Universal Coordenado), o que possibilitou a integração das duas bases com base na data e no horário das medições registradas. Baseando-se na data das medições, as variáveis foram organizadas de forma que as informações referentes ao mesmo dia fossem agregadas em uma única instância de dados. A partir da análise dos valores da variável Risco de Fogo presente na base do INPE, foi possível estabelecer uma classificação dos níveis de risco. Os rótulos definidos para essa variável foram feitos conforme a Tabela 1.

Tabela 1: Classificação de risco de fogo.

Classificação	Valor Correspondente
Baixo	Risco maior ou igual a 0 e menor que 0,4
Médio	Risco maior que 0,4 e menor que 0,7
Alto	Risco maior que 0,7 e menor ou igual a 1

Fonte: Elaborado pelos autores a partir da base de dados do INPE (2025).

2.3. Organização da Base de Dados

No intuito de evitar resultados tendenciosos à classe majoritária, no caso a classificação alto, que obteve mais do quádruplo de amostras que as outras classes, foi empregada uma técnica de balanceamento de dados híbrida. Em um estudo conduzido por Wongvorachan; He;

Bulut(2023), foi comprovado que, para dados extremamente desbalanceados, uma abordagem que combinasse o SMOTE-NC(Synthetic Minority Oversampling Technique for Nominal and Continuous) e RUS (Random Undersampling) é mais eficaz. Dessa forma, no presente trabalho, o SMOTE foi configurado para sintetizar pontos de dados minoritários com base em cinco vizinhos mais próximos e uma razão de reamostragem de 70% da classe majoritária, depois o RUS foi implementado para reduzir aleatoriamente o número de amostras da classe majoritária.

2.4. Algoritmo de Aprendizado Random Forest

O algoritmo Random Forest funciona como uma vasta “floresta” composta por múltiplas “árvores de decisão” individuais. Cada árvore é gerada de forma independente, utilizando uma seleção aleatória de dados ou características para embasar suas decisões, o que torna cada uma delas ligeiramente distinta. Ao realizar uma previsão, cada árvore da floresta oferece seu “palpite”. Para tarefas de classificação (previsão de categorias), o Random Forest adota a resposta indicada pela maioria das árvores. Já para problemas de regressão (previsão de valores numéricos), a previsão final corresponde à média dos resultados fornecidos por todas as árvores. Essa abordagem combinada torna o algoritmo altamente robusto e eficiente (BREIMAN, 2001).

Para a aplicação do algoritmo de aprendizado foi necessário normalizar conjuntamente todas as variáveis preditoras, uma vez que eram exclusivamente do tipo numérico. Por outro lado, a variável resposta não exigiu processo de normalização. O método adotado para normalização foi o de escala padrão, no qual as médias das distribuições são subtraídas dos dados e estes são redimensionados para apresentar variância unitária. O método de hiperparâmetro utilizado foi o de busca aleatória no qual foram testados 20 configurações diferentes dentro dos espaços definidos para cada hiperparâmetro, o resultado gerado foi que com conjunto de dados disponível o modelo teve um bom desempenho ao operar por meio do critério *log loss* e com uma floresta composta por 500 árvores, cada uma crescendo de forma relativamente diversa. Para promover diversidade entre as árvores e reduzir a correlação entre elas, cada divisão de nó deve considerar apenas a raiz quadrada do número total de variáveis, o que contribui para uma conjunto de modelos mais robustos.

2.5. Algoritmo de Aprendizado Extremely randomized trees

O Extra Trees também é um algoritmo que se baseia em uma “árvore de decisão”, a diferença é que ele divide os nós escolhendo pontos de corte de forma totalmente aleatória, sendo que as previsões das árvores são reunidas para produzir a previsão final (GEURTS et al, 2006). O método adotado para normalização foi novamente o de escala padrão, também repetiu-se o método de hiperparâmetro utilizado, o de busca aleatória, no qual foram testados 20 configurações diferentes dentro dos espaços definidos para cada hiperparâmetro. O resultado foi que com o conjunto de dados disponível, o modelo Extra Trees teve um bom desempenho ao utilizar o critério *entropy*, para ganho de informação, com uma floresta composta por 200 árvores, considerando até nove atributos para cada divisão de nó. Diferente do Random Forest, as árvores no Extra Trees não usam *bootstrap* (amostragem com reposição) e as divisões dos nós são escolhidas de forma totalmente aleatória.

2.6. Métricas de Avaliação

O objetivo das métricas de avaliação é comparar os algoritmos utilizados na análise dos dados e avaliar a eficácia de cada um deles. A partir do estudo dos valores da variável “RiscoFogo” do banco de dados de focos de incêndios do INPE, realiza-se a classificação dos níveis de risco de fogo, que são divididos em três categorias: baixo, médio e alto.

A avaliação dos algoritmos é feita por meio do processamento de bases de dados históricas, permitindo que eles estimem a quantidade de focos de incêndio em um certo período e a comparem com os registros já existentes. O valor previsto pelo algoritmo é comparado com o dado real, e sua performance é avaliada utilizando a métrica de acurácia.

2.6.1. Acurácia

A acurácia é utilizada para determinar a porcentagem de acertos, sendo calculada como a proporção entre o número total de acertos e o total de entradas analisadas (MARIANO, 2021).

$$\text{Acurácia} = \frac{\text{Total de acertos}}{\text{Total de entradas}}$$

2.6.2. Recall

O *recall*, também chamado de sensibilidade, mede a capacidade de um método em detectar adequadamente os casos classificados como positivos. Esse indicador é crucial para reduzir a ocorrência de falsos negativos, garantindo a identificação precisa dos verdadeiros positivos. Para obter seu valor, o número de verdadeiros positivos (VP) é dividido pelo total de verdadeiros positivos somados aos falsos negativos (FN) (MARIANO, 2021).

$$\text{Recall} = \frac{VP}{VP + FN}$$

2.6.3. Precisão

A precisão é um indicador que avalia a proporção de previsões corretas entre todas as classificações realizadas, mostrando a exatidão do modelo. Ela é calculada dividindo o número de verdadeiros positivos pelo total de todos os verdadeiros, isto é, tanto os verdadeiros positivos quanto os falsos positivos (FP) (MARIANO, 2021).

$$\text{Precisão} = \frac{VP}{VP + FP}$$

2.6.4 F-1 Score

O *F-1 Score* é um indicador frequentemente usado para avaliar a eficácia de um modelo de aprendizado de máquina. Essa medida é particularmente valiosa quando se deseja equilibrar *recall* e precisão, já que seu cálculo combina essas duas métricas em uma única avaliação (MARIANO, 2021).

$$f1 = 2 \times \frac{\text{precisão} \times \text{recall}}{\text{precisão} + \text{recall}}$$

2.6.5. Matriz de Confusão

A matriz de confusão serve como um instrumento eficaz para avaliar a performance de classificação. É uma tabela que compara quantitativamente as ocorrências reais de cada categoria com as previsões realizadas pelo modelo. A matriz de confusão é uma ferramenta estatística amplamente utilizada na avaliação da performance de classificadores, especialmente em estudos de sensoriamento remoto e mapeamento temático. Ela organiza os resultados da classificação em uma tabela onde as linhas representam as classes reais e as colunas representam as classes previstas pelo modelo, permitindo identificar tanto os acertos quanto os erros cometidos pelo algoritmo (STEHMAN *et al.*, 2024).

Essa estrutura possibilita a derivação de diversas métricas de desempenho, como a acurácia do produtor (relacionada ao erro de omissão), a acurácia do usuário (relacionada ao erro de comissão), além da acurácia global do modelo. A análise da matriz é particularmente útil em situações com desequilíbrio entre classes, pois permite observar se classes minoritárias estão sendo corretamente identificadas ou confundidas com classes mais frequentes. Segundo Stehman *et al.* (2024), o uso de matrizes de confusão com base em amostragem probabilística é essencial para garantir rigor estatístico na avaliação da precisão de mapas temáticos, além de possibilitar estimativas consistentes sobre o desempenho do modelo em toda a população mapeada.

3. Resultados/Discussões

Após a mesclagem das bases de dados, foram obtidas 105.921 amostras, das quais 12.233 pertencem à classe de risco de fogo baixo, 18509 à classe de risco de fogo médio e 75.179 à classe de risco alta. foi observada diferença significativa para todas as classes de risco de fogo(baixo, médio e alto), considerando o p-valor menor que 0,05. Todas as variáveis consideradas apresentaram diferença significativa, menos FRP, a qual foi retirada da base de dados para treinamento. Além disso, devido ao alto desbalanceamento de dados entre as classes, foi aplicada a técnica híbrida de balanceamento de dados, conforme indicada na metodologia, o que resultou em um total de 157.875 amostras, com 52.625 pontos de dados para cada classe.

A classificação dos focos de incêndio em Altamira, região com histórico crítico, utilizou um modelo de machine learning baseado no algoritmo Random Forest e Extra Trees. A Tabela 2 expõe de forma detalhada os resultados de cada modelo .

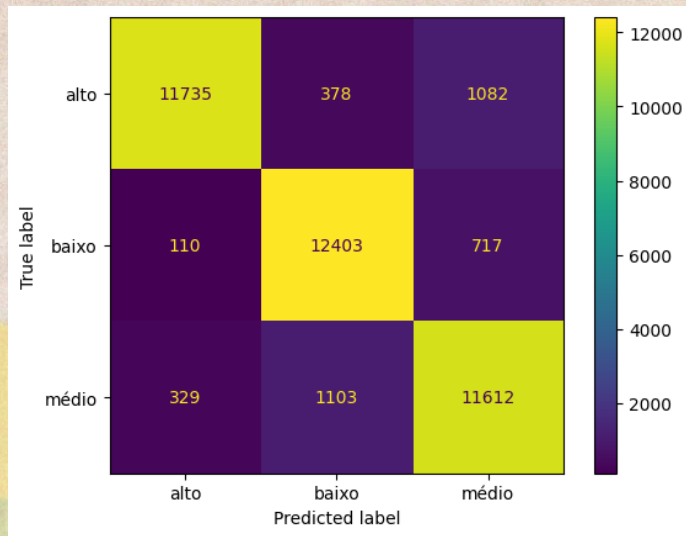
Tabela 2: Resultados dos Algoritmos.

Random Forest				
Classificação	Acurácia	Precisão	Recall	F1-score
Alto	90,4%	96%	89%	92,6%
Baixo		89,2%	93,2%	91%
Médio		86,2%	89,2%	87,8%
Extra Trees				
Classificação	Acurácia	Precisão	Recall	F1-score
Alto	90,8%	96%	89,2%	92,6%
Baixo		89,6%	93,6%	91,6%
Médio		86,8%	89,2%	88%

Fonte: Elaborado pelos autores (2025).

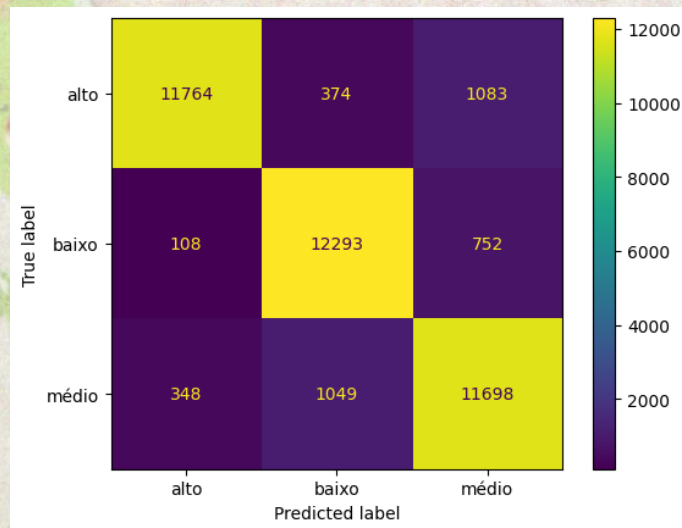
As figuras 1 e 2 dizem respeito às matrizes confusão do Random Forest e Extra Trees, respectivamente.

Figura 1 - Matriz confusão do modelo Random Forest



Fonte: Elaborado pelos autores (2025).

Figura 2 - Matriz confusão do modelo Extra Trees



Fonte: Elaborado pelos autores (2025).

A partir dos resultados obtidos, percebe-se uma ligeira vantagem do Extra Trees, com uma acurácia de 90,8%, sobre a acurácia de 90,4% do Random Forest. Ademais, os algoritmos demonstram resultados satisfatórios nas demais métricas, em específico, a precisão de 96% que ambos algoritmos atingiram ao identificarem casos com alto risco de incêndio. Além da precisão, a classificação de risco de fogo alto destacou-se na métrica f1-score nos dois modelos, com 92,6%. Já na métrica recall, a classificação de risco de fogo baixo demonstrou maior eficiência de predição dentre as classes, com 93,2%.

Somente a classe de risco de fogo médio não alcançou acima de 90% em nenhuma métrica, embora tenha superado os 85% em todas. Tais resultados menos precisos se devem, provavelmente, ao seu aspecto intermediário, ou seja, os valores das features nesta classe assumem um papel “mediador” entre as classes alto e baixo, causando certa dificuldade aos modelos para prevê-la. Como percebido nas matrizes de confusão dos modelos, esse rótulo obteve a menor quantidade de acertos, sendo confundido com a classe de risco de fogo baixo mais vezes que a outra classe. Enquanto isso, os outros rótulos foram mais comumente confundidos com a classe de risco de fogo médio.

Outrossim, a seleção aleatória de dados é uma característica comum entre os algoritmos Random Forest e Extra Trees (BREIMAN, 2001; GEURTS *et al*, 2006), o que demonstra que o conjunto de dados utilizado no treinamento se adequa de forma satisfatória a essa natureza. Essa afirmação também se apoia na superior acurácia do Extra Trees, cuja aleatoriedade está presente tanto na seleção de atributos quanto nos pontos de corte em cada nó, além do uso da amostra completa de aprendizado, em vez de réplicas de *bootstrap* (GEURTS *et al*, 2006).

4. Conclusão

O estudo mostrou que os algoritmos Random Forest e Extra Trees são eficazes na classificação do risco de incêndios em Altamira, com a acurácia de 90,4% e 90,8%, respectivamente. Os modelos destacam-se na previsão de focos de alto e baixo risco, mas teve desempenho inferior na classe de risco médio, o que sugere a necessidade de melhorias na abordagem. Técnicas de balanceamento de dados, como SMOTE e RUS, foram essenciais para lidar com o desbalanceamento entre as classes de risco. Embora os resultados sejam promissores, expandir a base de dados e incorporar variáveis adicionais pode aprimorar ainda mais a precisão do modelo. A adaptação dos algoritmos à outras regiões de risco mostra seu grande potencial como ferramenta de previsão e monitoramento de incêndios, podendo ser integrado a sistemas de alerta precoce para mitigar danos ambientais e humanos.

5. Referências Bibliográficas

BREIMAN, L. Random Forests. *Machine Learning*, v. 45, n. 1, p. 5–32, out. 2001.

CHANG, Jaesung; YOON, Jaeyoung; LEE, Gunho. Machine Learning Techniques in Structural Fire Risk Prediction. *International Journal of Software Engineering and Its Applications*, v.14, p. 17-26, Jun. 2020.

GALVAO JUNIOR, P. A.; ROVEDA, S. R. M. M.; VIEIRA, H. E. DE M. Modelos Híbridos Aplicados à Construção de Índice de Classificação de Níveis de Risco de Fogo no Brasil. *Revista Brasileira de Ciências Ambientais*, v. 57, n. 3, p. 364–374, 26 ago. 2022.

GEURTS, Pierre; ERNST, Damien; WEHENKEL, Louis. Extremely randomized trees. *Machine Learning*, v.63, p.3-42, Mar 2006.

INPE. Situação atual. 2025. Disponível em:

https://terrabrasilis.dpi.inpe.br/queimadas/situacao-atual/situacao_atual/. Acesso em: 13 jun. 2024.

MANIATIS, Y.; DOGANIS, A.; CHATZIGEORGIADIS, M. Fire Risk Probability Mapping Using Machine Learning Tools and Multi-Criteria Decision Analysis in the GIS Environment: A Case Study in the National Park Forest Dadia-Lefkimi-Soufli, Greece. *Applied Sciences*, v. 12, n. 6, p. 2938, 13 mar. 2022.

MAPBIOMAS BRASIL. Disponível em: <https://brasil.mapbiomas.org/2025/01/22/area-queimada-no-brasil-cresce-79-em-2024-e-supera-os-30-milhoes-de-hectares/>. Acesso em: 13 jun. 2025.

MARIANO, D. Métricas de avaliação em machine learning: acurácia, sensibilidade, precisão, especificidade e F-score. *BIOINFO - Revista Brasileira de Bioinformática e Biologia Computacional*, jul. 2021.

MARTIN, D.; TOMIDA, M.; MEACHAM, B. Environmental impact of fire. *Fire Science Reviews*, v. 5, n. 1, p. 5, 2016.

STEHMAN, S. V. et al. A population-based confusion matrix approach for accuracy assessment of map classifications. *Remote Sensing*, [S. l.], v. 16, n. 3, p. 533, 2024.

Disponível em: <https://www.mdpi.com/2072-4292/16/3/533>. Acesso em: 20 jun. 2025.

SHAO, Y. et al. Mapping China's Forest Fire Risks with Machine Learning. *Forests*, v. 13, n. 6, p. 856, 30 maio 2022.

TAN, C.; FENG, Z. Mapping Forest Fire Risk Zones Using Machine Learning Algorithms in Hunan Province, China. *Sustainability*, v. 15, n. 7, p. 6292, 6 abr. 2023.

WONGVORACHAN, T.; HE, S.; BULUT, O. A Comparison of Undersampling, Oversampling, and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining. *Information*, v. 14, n. 1, p. 54, 16 jan. 2023.

