

XXVI ENMC
ENCONTRO NACIONAL DE MODELAGEM COMPUTACIONAL

XIV ECTM
ENCONTRO DE CIÊNCIA E TECNOLOGIA DE MATERIAIS



AVALIAÇÃO EXPLORATÓRIA DO ALGORITMO RANDOM FOREST PARA PREVISÃO DE INCÊNDIOS FLORESTAIS

Kemuel Maciel Freitas^{1,2} – kemuel.macielf@gmail.com

Carlos Henrique Pinto da Silveira³ – carlos.silveira@discente.ufopa.edu.br

Christiano Jorge Gomes Pinheiro⁴ – christiano.pinheiro@ufes.br

Jefferson Rossy Pereira da Silva^{1,2} – j.rossy95@gmail.com

Ronie Silva Juvanhol^{1,5} – roniejuvanhol@ufpi.edu.br

Anderson Alvarenga de Moura Meneses^{1,2} – anderson.meneses@ufopa.edu.br

¹ Programa de Pós-Graduação em Recursos Naturais da Amazônia, Instituto de Engenharia e Geociências, Universidade Federal do Oeste do Pará - Santarém, PA, Brasil

² Laboratório de Inteligência Computacional, Instituto de Engenharia e Geociências, Universidade Federal do Oeste do Pará - Santarém, PA, Brasil

³ Laboratório de Modelagem Computacional, Instituto de Engenharia e Geociências, Universidade Federal do Oeste do Pará – Santarém, PA, Brasil

⁴ Departamento de Engenharia Rural, Universidade Federal do Espírito Santo – Alegre, ES, Brasil

⁵ Departamento de Engenharia Florestal, Universidade Federal do Piauí – Bom Jesus, PI, Brasil

Resumo. *O objetivo deste trabalho é avaliar o desempenho do algoritmo Random Forest (RF) para a previsão da susceptibilidade de incêndios florestais por meio de um estudo de caso utilizando um conjunto de dados previamente validados e amplamente aceitos na comunidade científica. Para isso foi utilizado um conjunto de dados obtidos no Estado do Espírito Santo, tendo como variável resposta a Densidade Kernel calculada sobre 3314 pontos de queima no referido estado, e mais 12 variáveis preditoras: Altitude, Declividade, Índice Topográfico Composto, Precipitação Média Anual, Temperatura Média Anual, Radiação Solar, Deficiência Hídrica Média Anual, Densidade Demográfica, Renda, Proximidade a Estradas, Campo Contínuo de Vegetação e Uso e Cobertura da Terra. O desempenho do RF foi comparado com resultados já validados para Árvores de Classificação e Regressão (CART) por meio do coeficiente de determinação (R^2) e da raiz do erro quadrático médio (RMSE). O RF apresentou resultados ligeiramente melhores em ambas as métricas, com um R^2 de 0,9718 em comparação aos 0,88 do CART, e RMSE de 0,0162 contra 0,55, sugerindo maior capacidade de capturar complexidades nos dados e gerar previsões mais precisas.*

Palavras-Chave: *Sensoriamento Remoto, Aprendizagem de Máquina, Random Forest, Previsão de Incêndios Florestais*

1. INTRODUÇÃO

O uso do fogo é uma prática comum para a limpeza de áreas no Brasil, principalmente de floresta, que após o corte e a queima é transformada em pastagem, em que inicialmente as cinzas fornecem nutrientes ao solo sob pastagem, tornando a queimada uma opção rápida e barata, mesmo que após algum tempo seja necessária a adubação (Alves & Modesto Junior, 2020). Entretanto, se realizada de maneira imprudente e, até criminosa, essa atividade pode ocasionar uma série de prejuízos ambientais, econômicos e sociais, além de trazer riscos à vida humana (Fonseca-Morello et al. 2017).

Inicialmente, é necessário entender por completo os processos e interações que levam à ignição e a propagação do fogo, para que, então possam ser desenvolvidas ações preventivas e de controle da ocorrência, da intensidade e da magnitude das queimadas (Mohajane et al. 2021; Oliveira et al. 2020). O comportamento e as características do fogo são determinados por uma série de fatores, entre eles relevo, vegetação e clima, tornando uma tarefa complexa prever em quais áreas devem-se alocar os recursos para combate e prevenção dos incêndios florestais

Para isso, diversos modelos que utilizam métodos de *machine learning* e *deep learning* já têm sido desenvolvidos para predição de incêndios florestais em diferentes contextos e situações, utilizando Redes Neurais, Algoritmos de Árvores de Regressão e Classificação, Random Forest, Máquina de Vetores de Suporte, Redes Neurais Profundas, XGBoost, além de vários outros, na Ásia (Pourghasemi et al. 2020; Pang et al. 2022), na Europa (Tonini et al. 2020; Michael et al. 2021), na América do Norte (Phelps e Woolford, 2021; Abid, 2021) e também no Brasil, principalmente nas regiões sudeste e centro-oeste (Juvanhol et al. 2023; Rubí et al. 2023).

Contudo, a escolha do modelo mais adequado para essa tarefa não é uma escolha trivial e levanta uma série de questionamentos em relação à escolha correta diante da ampla gama de algoritmos disponíveis. A problemática central reside em determinar qual modelo maximiza a precisão das previsões, considerando a natureza dinâmica e complexa dos sistemas envolvendo os incêndios florestais. Dessa forma, o objetivo deste trabalho é avaliar o desempenho do algoritmo Random Forest para a previsão da susceptibilidade de incêndios florestais por meio de um estudo de caso utilizando um conjunto de dados previamente validados e amplamente aceitos na comunidade científica, visando garantir a comparação justa e imparcial do desempenho do algoritmo em relação a outras abordagens e modelos.

2. REFERENCIAL BIBLIOGRÁFICO

2.1 Inteligência Artificial

A inteligência artificial surgiu entre as décadas de 1940 e 1950, de modo que a primeira visão geral foi apresentada por Alan Turing. O teste de Turing se baseia no conceito de inteligência, de modo que um computador passará no teste se um interrogador humano não for capaz de distinguir entre as respostas do computador e de um outro humano (Turing, 2009). Os modelos de IA foram propostos a partir de dois objetivos, o primeiro é o de desenvolver sistemas ou entidades que realizem tarefas que só poderiam ser realizadas por humanos ou aquelas tarefas que não poderiam ser resolvidas a partir de uma solução algorítmica computacional convencional (Norvig e Russel, 2013).

Em 1956, John McCarthy elaborou o primeiro conceito de Inteligência Artificial como “a capacidade de uma máquina realizar funções que, se realizadas por um ser humano, seriam consideradas inteligentes” (Mccarthy, 2004). Segundo Haykin (1998), sistemas de Inteligência Artificial devem ser capazes de realizar três coisas: Armazenar conhecimento; aplicar esse conhecimento armazenado à resolução de problemas e adquirir novos conhecimentos por meio das suas experiências.

De acordo com Costafreda-Aumedes *et al.* (2017), os primeiros modelos de inteligência artificial para ocorrência de fogo surgiram na segunda metade dos anos 1980, tendo evoluído ao longo dos anos acompanhando a evolução do acesso à computadores cada vez mais potentes e do acesso à quantidades cada vez maiores de dados, chegando aos modelos utilizados atualmente, como Árvores de Classificação e Regressão – CART; Redes Neurais Artificiais – RNA; Máquina de Vetores de Suporte – SVM, Random Forest – RF, etc.

2.2 Árvore de Classificação e Regressão (CART)

O algoritmo de Árvore de Classificação e Regressão (*Classification and Regression Tree* – CART) foi proposto inicialmente por Breiman *et al.* (1984) e desde então, tem sido amplamente utilizado em diversos campos, incluindo análise de dados, aprendizado de máquina e previsão de eventos complexos, como incêndios florestais.

As Árvores de Decisão, como as implementadas no algoritmo CART, são modelos que buscam segmentar os dados em subgrupos homogêneos com base em critérios de divisão, o processo começa com um nó raiz que representa todo o conjunto de dados (Lozano *et al.* 2008). O algoritmo seleciona, de maneira iterativa, a variável e o ponto de corte que melhor separam os dados em dois grupos distintos, otimizando a métrica selecionada (Mohajane *et al.* 2021), como por exemplo, a raiz do erro médio quadrático (*Root Mean Squared Error* – RMSE).

A cada divisão, novos nós filhos são criados, representando subgrupos mais específicos dos dados. Esse processo é repetido recursivamente até que algum critério de parada seja alcançado, como um tamanho mínimo de amostra em um nó ou a profundidade máxima da árvore. Uma vez que a árvore é construída, ela pode ser usada para fazer previsões. Quando novos dados são apresentados, eles percorrem a árvore, seguindo os caminhos determinados pelos critérios de divisão, e a previsão é feita com base nos valores médios ou na classe majoritária do grupo terminal correspondente.

A simplicidade e interpretabilidade das Árvores de Decisão, como implementadas pelo algoritmo CART, são características valorizadas. A estrutura da árvore permite visualizar e entender as decisões tomadas pelo modelo, tornando-o atraente para situações em que a transparência é essencial (Mohajane *et al.* 2021). No entanto, o CART pode ser suscetível ao *overfitting* (Rasekhschaffe & Jones, 2019), especialmente quando a árvore cresce muito e se ajusta aos ruídos dos dados de treinamento. Essa complexidade pode ser limitada por meio de técnicas como a poda da árvore ou a definição de critérios de parada adequados.

2.3 Random Forest (RF)

O algoritmo Random Forest é uma evolução das Árvores de Decisão e busca superar algumas de suas limitações, sendo proposto por Breiman (2001). Ele pertence à categoria de algoritmos *ensemble*, que combinam as previsões de múltiplos modelos para melhorar a robustez e a precisão das previsões (Da Silva *et al.* 2022).

O funcionamento do Random Forest é baseado na criação de uma coleção de Árvores de Decisão, cada uma construída a partir de diferentes subconjuntos de dados de treinamento e variáveis aleatórias. Essa abordagem de aleatoriedade visa mitigar o *overfitting*, um problema comum nas Árvores de Decisão tradicionais. Durante a construção de cada árvore, em cada divisão de nó, apenas um subconjunto das variáveis é considerado para seleção do melhor critério de divisão. Isso reduz a tendência de uma árvore superajustar-se aos dados de treinamento (Mohajane *et al.* 2021).

Quando se trata de fazer previsões, cada árvore individual do Random Forest emite uma previsão e a classe ou valor resultante é escolhido por meio da média (no caso da regressão) entre as previsões de todas as árvores (Iannace *et al.* 2019). Esse processo de combinar

previsões de várias árvores reduz o impacto de anomalias e ruídos nos dados, resultando em previsões mais robustas e precisas (Breiman, 2001).

O Random Forest oferece diversas vantagens, incluindo a capacidade de lidar bem com dados desbalanceados, robustez a outliers e uma relativa resistência ao overfitting (Theodoris & Tsadiras, 2022). Além disso, ao contrário das Árvores de Decisão, o Random Forest não requer ajustes complexos de hiperparâmetros, o que facilita seu uso prático.

3. MATERIAL E MÉTODOS

As principais etapas metodológicas do estudo estão esquematizadas na Fig. 1. Os dados foram obtidos a partir de diversas bases de dados, e após o cálculo da densidade kernel, todas as variáveis foram transformadas em imagens raster com resolução espacial de 250 metros. Em seguida, as imagens foram carregadas em ambiente de Sistema de Informações Geográficas (SIG), utilizando o software *QGIS 3.28.7*, onde foram amostrados os valores de cada variável e aglomerados em uma única tabela de atributos, gerando um arquivo com formato CSV.

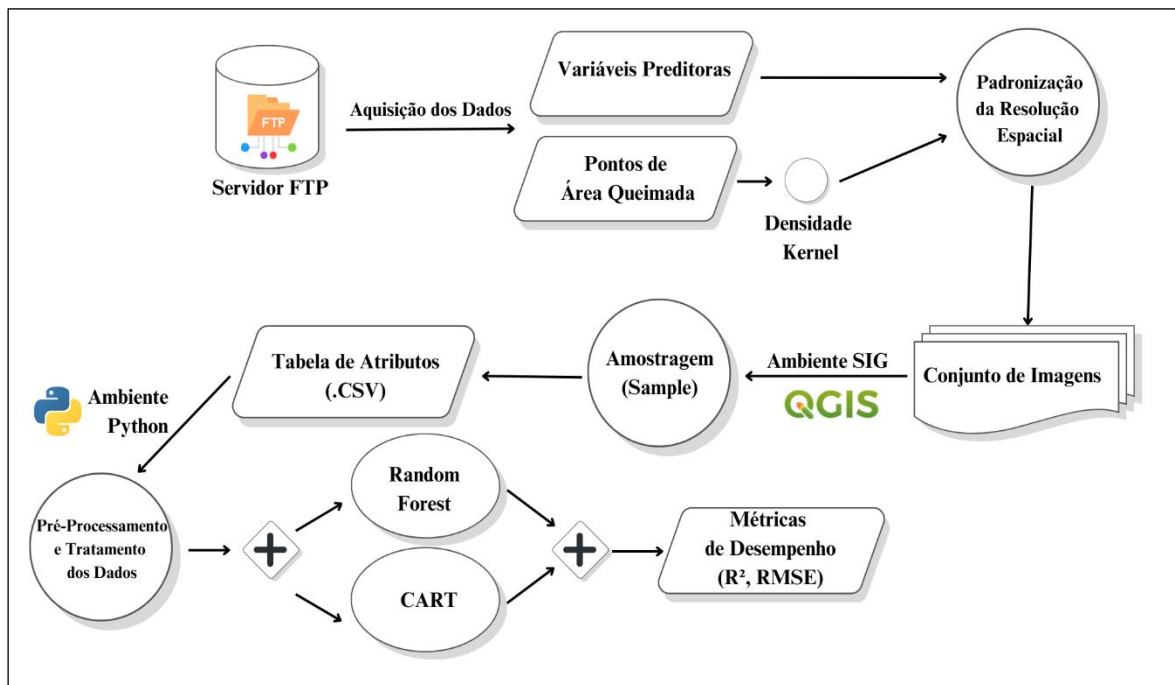


Figura 1: Fluxograma dos principais procedimentos e ferramentas envolvidos
Fonte: Autor (2023)

3.1 Estudo de Caso

Para este estudo de caso, foram utilizados os dados apresentados no artigo “*Use of machine learning as a tool for determining fire management units in the brazilian atlantic forest*” (Juvanhol *et al.* 2023), que teve por objetivo avaliar a predição de fogo em escala regional e fornecer saídas compreensíveis, na forma de regras de decisão, capazes de prever os valores de risco médio para cada célula de grade, definindo assim as unidades de manejo do fogo no estado do Espírito Santo.

O dataset do estudo abrange dados do período entre os anos de 2000 e 2015, contendo uma variável-alvo, representada pela Densidade Kernel calculada sobre 3314 pontos de área de queima; além de 12 variáveis preditoras: Altitude, Declividade, Índice Topográfico Composto, Precipitação Média Anual, Temperatura Média Anual, Radiação Solar, Deficiência Hídrica

Média Anual, Densidade Demográfica, Renda, Proximidade a Estradas, Campo Contínuo de Vegetação e Uso e Cobertura da Terra.

3.2 Configurações de Software e Hardware

A partir da aquisição do dataset com os dados da pesquisa de Juvanhol *et al.* (2023), os dados foram inseridos diretamente em ambiente de programação Python, onde foram submetidos a um novo modelo, o Random Forest, a partir de validação cruzada com $k = 10$. Finalmente, as médias das 10 execuções de cada uma das métricas foram comparadas às obtidas por Juvanhol *et al.* (2023) ao utilizarem o algoritmo de Árvore de Classificação e Regressão (CART – Classification And Regression Tree). A RF foi implementada com 100 estimadores (parâmetro $n_estimators$), a partir da classe *RandomForestRegressor*, do módulo *ensemble*, da biblioteca *scikit-learn 1.3.0*.

Para realizar o treinamento e teste, um computador com as seguintes configurações foi utilizado: processador Intel® Core™ i5 - 7400, 16GB de RAM e GPU Nvidia RTX 2080Ti com 11GB de VRAM.

3.3 Avaliação dos Modelos

Para verificar a capacidade de cada um dos modelos de prever os valores da variável resposta foram avaliadas as métricas *R-Squared* – R^2 e *Root Mean Squared Error* – RMSE, dadas pelas equações 1 e 2:

$$R^2(x, y) = \frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (1)$$

$$RMSE(x, y) = \sqrt{\frac{1}{n} \sum_{i=1}^n (x - y)^2} \quad (2)$$

nas quais x representa os valores reais observados e y representa os valores preditos. Os resultados de R^2 vão de 0 a 1, de forma que, quanto mais próximo de 1, mais explicativo é o modelo em relação aos dados previstos. Diferente do R^2 , os valores do RMSE não possuem um limite máximo e medem a diferença entre os valores reais e os valores preditos e acabam penalizando possíveis valores discrepantes, os chamados *outliers* (Harrison, 2020). Vale ressaltar que diferente da métrica R^2 , na qual os resultados mais próximos de um são considerados “melhores”, para a métrica de erro (RMSE), quanto menores forem os valores, melhores serão os resultados.

4. RESULTADOS E DISCUSSÃO

Ao submeter o conjunto de dados ao algoritmo CART, Juvanhol *et al.* (2023) obtiveram um R^2 médio de 0,88, valor relativamente próximo ao observado nos testes realizados com o Random Forest. Por outro lado, o algoritmo CART apresentou um Erro Quadrático Médio (MSE) médio de 0,01 com uma configuração de árvore com 5137 nós terminais, que após o procedimento de poda, chegou a um MSE de 0,3 em uma configuração com 38 nós terminais.

Os valores de MSE podem ser convertidos para RMSE apenas aplicando uma raiz quadrada aos valores, ou seja, o valor final do RMSE para o algoritmo CART ficaria 0,55. Desta forma, pode-se observar um melhor desempenho do algoritmo RF com relação às métricas. Porém uma desvantagem deste algoritmo é que não é possível conhecer as estruturas de cálculo internas, o que é possível com o algoritmo CART utilizado por Juvanhol *et al.* (2023).

Tabela 1 - Comparação do Resultado médio das métricas obtidas a partir da validação cruzada com os resultados obtidos por Juvanhol et al. (2023) para o mesmo conjunto de dados

Algoritmo	Métrica	
	R ²	RMSE
RF – Random Forest	0,9718	0,0162
CART – Classification and Regression Tree*	0,88	0,55

* Resultados obtidos por Juvanhol et al. (2023)

Fonte: Autor (2023)

Ao comparar as pontuações de influência de cada variável nos modelos CART e RF (Tabela 2) não existem grandes variações, exceto pelo Uso e Cobertura da Terra que no CART teve uma pontuação de 75,07%, enquanto no RF, apenas 1,89%. Contudo, no geral, os fatores socioeconômicos e ambientais continuam tendo maior importância no RF, assim como no CART, mantendo a tendência apontada por Juvanhol *et al.* (2023), da maior importância desses fatores em detrimento de menores valores de importância para os fatores topográficos, e também da pouca significância das variáveis campo contínuo e radiação.

Tabela 2 - Comparação em termos percentuais das Pontuações de cada variável para os modelos CART e Random Forest

VARIÁVEL	PONTUAÇÃO (%)	
	RF	CART*
Densidade Demográfica	100	100
Precipitação Média Anual	86,13	78,4
Renda	42,23	36,05
Altitude	35,69	33,51
Temperatura Média Anual	32,79	22,13
Deficiência Hídrica Média Anual	23,47	22,89
Índice Topográfico Composto	3,79	24,06
Proximidade a Estradas	2,14	9,92
Campo Contínuo de Vegetação	2,03	1,04
Uso e Cobertura da Terra	1,89	75,07
Declividade	0,94	9,14
Radiação Solar	0,39	0,01

* Resultados obtidos por Juvanhol *et al.* (2023)

Fonte: Autor (2023)

Apesar de no contexto deste estudo, onde o desempenho da RF demonstrou ser ligeiramente superior com base nas métricas R² e RMSE, a decisão sobre qual entre essas duas abordagens é mais indicada para o problema proposto deve considerar diversos fatores. Primeiramente, a diferença nas métricas R² e RMSE indica que a RF está apresentando uma melhor capacidade de explicar a variabilidade dos dados e de reduzir o erro de previsão em comparação com CART.

Além disso, é importante considerar a complexidade dos modelos. As Árvores de Classificação e Regressão (CART), tendem a ser mais interpretáveis, uma vez que podem ser visualizadas e compreendidas de maneira mais direta (Mohajane *et al.* 2021). Isso pode ser crucial para áreas em que a interpretabilidade do modelo é tão importante quanto a precisão da previsão, permitindo que especialistas e tomadores de decisão entendam as razões por trás das previsões feitas pelo modelo. Por outro lado, a RF é uma abordagem mais complexa que envolve a combinação de várias árvores (Iannace *et al.* 2019), o que pode dificultar a interpretação dos resultados.

Outro aspecto relevante é a escalabilidade. A Random Forest, por ser composta por várias árvores, pode demandar mais recursos computacionais, como memória e poder de processamento, em comparação com CART. Isso deve ser levado em consideração, especialmente se o modelo for implementado em um ambiente com recursos limitados.

Além disso, se houver interesse em compreender como o modelo toma suas decisões, CART pode fornecer uma visão mais clara das regras e caminhos de decisão adotados. Isso pode ser fundamental em situações em que a transparência é um requisito crucial para a adoção do modelo.

Considerando a interpretabilidade, as Árvores de Decisão CART mostraram-se mais acessíveis, permitindo uma compreensão clara das regras de decisão adotadas pelo modelo. Essa característica pode ser vital em cenários em que a explicação das previsões é essencial para engajar especialistas e tomadores de decisão. Por outro lado, a RF, embora mais complexa, forneceu um desempenho superior, o que a torna uma escolha atraente em casos onde a máxima precisão é o objetivo principal.

5. CONCLUSÕES

Este estudo ressalta a importância de avaliar cuidadosamente os *trade-offs* entre precisão, interpretabilidade e recursos computacionais ao selecionar um modelo para a previsão da susceptibilidade a incêndios florestais. A compreensão das estruturas de cálculo internas de cada algoritmo pode fornecer insights adicionais para a tomada de decisão. Nesse contexto, a RF demonstrou um potencial promissor para melhorar a previsão da susceptibilidade a incêndios, embora seja necessário ponderar esses benefícios em relação aos requisitos específicos de cada aplicação.

Acknowledgements

Os autores agradecem o apoio financeiro concedido pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e pela Universidade Federal do Oeste do Pará (UFOPA), aos Laboratórios de Inteligência Computacional (LABIC) e de Modelagem Computacional (LABMC) da UFOPA pelo acesso aos equipamentos e ao Instituto de Engenharia e Geociências (IEG) pelo apoio à ida dos congressistas ao evento.

REFERÊNCIAS

- Abid, F. (2021), A survey of machine learning algorithms based forest fires prediction and detection systems. *Fire technology*, v. 57, n. 2, p. 559-590. <https://doi.org/10.1007/s10694-020-01056-z>
- Alves, R.N.B.; Modesto Junior, M.S. (2020), “*Roça sem fogo: da tradição das queimadas para a agricultura sustentável na Amazônia*”. Brasília: Embrapa. 184 p.
- Breiman, L.; Friedman, J. H.; Olshen, R. A.; Stone, C. J. (1984), Classification and regression trees. Brooks. Wadsworth and Brooks, Monterey, CA. <https://doi.org/10.1201/9781315139470>

- Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32. <https://doi.org/10.1023/A:1010933404324>
- Costafreda-Aumedes, S.; Comas, C.; Vega-Garcia, C. (2017), Human-caused fire occurrence modelling in perspective: A review. *International Journal of Wildland Fire*, 26(12), 983-998. <https://doi.org/10.1071/WF17026>
- Da Silva, D. G.; Geller, M. T. B.; dos Santos Moura, M. S.; de Moura Meneses, A. A. (2022), Performance evaluation of LSTM neural networks for consumption prediction. *e-Prime-Advances in Electrical Engineering, Electronics and Energy*, 2, 100030. <https://doi.org/10.1016/j.prime.2022.100030>
- Fonseca-Morello, T.; Ramos, R.; Steil, L.; Parry, L.; Barlow, J.; Markusson, N. and Ferreira, A. (2017), Queimadas E Incêndios Florestais Na Amazônia Brasileira: Porque As Políticas Públicas Têm Efeito Limitado? *Ambiente & Sociedade*, v. 20, p. 19-38. <https://doi.org/10.1590/1809-4422asoc0232r1v2042017>
- Harrison, M. (2020) “*Machine Learning: Guia de referência rápido*”. 1ª edição. São Paulo — SP: Novatec Editora Ltda.
- Haykin, S. (1998), *Neural networks: a comprehensive foundation*. Prentice Hall PTR.
- Iannace, G.; Ciaburro, G.; Trematerra, A. (2019), Wind turbine noise prediction using random forest regression. *Machines*, 7(4), 69. <https://doi.org/10.3390/machines7040069>
- Juvanhol, R. S.; Fiedler, N. C.; SANTOS, A. R.; Peluzio, T. M.; SILVA, W. B.; Pinheiro, C. J. G.; SOUSA, H. C. P. (2023), Use of machine learning as a tool for determining fire management units in the Brazilian Atlantic forest. *Anais da Academia Brasileira de Ciências*, v. 95, p. e20201039. <https://doi.org/10.1590/0001-3765202320201039>
- Lozano, F. J.; Suárez-Seoane, S.; Kelly, M.; Luis, E. (2008), A multi-scale approach for modeling fire occurrence probability using satellite data and classification trees: A case study in a mountainous Mediterranean region. *Remote Sensing of Environment*, 112(3), 708-719. <https://doi.org/10.1016/j.rse.2007.06.006>
- McCarthy, J. (2007), What is artificial intelligence. Disponível em: <http://www-formal.stanford.edu/jmc/whatisai.pdf>
- Michael, Y.; Helman, D.; Glickman, O.; Gabay, D.; Brenner, S.; Lensky, I. M. (2021), Forecasting fire risk with machine learning and dynamic information derived from satellite vegetation index time-series. *Science of The Total Environment*, v. 764, p. 142844. <https://doi.org/10.1016/j.scitotenv.2020.142844>
- Mohajane, M.; Costache, R.; Karimi, F.; Pham, Q. B.; Essahlaoui, A.; Nguyen, H.; Laneve, G.; Oudija, F. (2021), Application of remote sensing and machine learning algorithms for forest fire mapping in a Mediterranean area. *Ecological Indicators*, v. 129, p. 107869. <https://doi.org/10.1016/j.ecolind.2021.107869>
- Norvig, P.; Russell, S. (2013), Inteligência artificial. *Rio de Janeiro: Grupo GEN*.
- Oliveira, V. F. R.; Da Silva, E. R. S.; Da Silva, B. H. M.; Vick, E. P.; Lima, C. G. R.; Bacani, V. M. (2020), Geoprocessamento aplicado ao mapeamento de risco a incêndios. *Revista Brasileira de Geografia Física*, v. 13, n. 03, p. 1194-1212. <https://doi.org/10.26848/rbgf.v13.3.p1194-1212>
- Pang, Y.; Li, Y.; Feng, Z.; Feng, Z.; Zhao, Z.; Chen, S.; Zhang, H. (2022), Forest fire occurrence prediction in China based on machine learning methods. *Remote Sensing*, v. 14, n. 21, p. 5546. <https://doi.org/10.3390/rs14215546>
- Phelps, N.; Woolford, D. G. (2021), Comparing calibrated statistical and machine learning methods for wildland fire occurrence prediction: A case study of human-caused fires in Lac La Biche, Alberta, Canada. *International journal of wildland fire*, v. 30, n. 11, p. 850-870. <https://doi.org/10.1071/WF20139>
- Pourghasemi, H. R.; Kariminejad, N.; Amiri, M.; Edalat, M.; Zarafshar, M.; Blaschke, T.; Cerda, A. (2020), Assessing and mapping multi-hazard risk susceptibility using a machine

- learning technique. *Scientific reports*, v. 10, n. 1, p. 3203. <https://doi.org/10.1038/s41598-020-60191-3>
- Rasekhschaffe, K. C.; Jones, R. C. (2019), Machine learning for stock selection. *Financial Analysts Journal*, 75(3), 70-88. <https://doi.org/10.1080/0015198X.2019.1596678>
- Rubí, J. N. S.; De Carvalho, P. H. P.; Gondim, P. R.L. (2023), Application of machine learning models in the behavioral study of forest fires in the Brazilian Federal District region. **Engineering Applications of Artificial Intelligence**, v. 118, p. 105649. <https://doi.org/10.1016/j.engappai.2022.105649>
- Theodoridis, G.; Tsadiras, A. (2022), Applying machine learning techniques to predict and explain subscriber churn of an online drug information platform. *Neural Computing and Applications*, 34(22), 19501-19514. <https://doi.org/10.1007/s00521-022-07603-9>
- Tonini, M.; D'Andrea, M.; Biondi, G.; Degli Esposti, S.; Trucchia, A.; Fiorucci, P. (2020), A machine learning-based approach for wildfire susceptibility mapping. The case study of the Liguria region in Italy. *Geosciences*, v. 10, n. 3, p. 105. <https://doi.org/10.3390/geosciences10030105>
- Turing, A. M. (2009). *Computing machinery and intelligence*. Springer Netherlands. https://doi.org/10.1007/978-1-4020-6710-5_3

EXPLORATORY EVALUATION OF THE RANDOM FOREST ALGORITHM FOR FOREST FIRE PREDICTION

Abstract. *The objective of this study is to evaluate the performance of the Random Forest (RF) algorithm for predicting forest fire susceptibility through a case study using a dataset previously validated and widely accepted in the scientific community. To achieve this, a dataset collected in the State of Espírito Santo was utilized, with the response variable being the Kernel Density calculated over 3314 fire ignition points in that state, along with 12 predictor variables: Altitude, Slope, Composite Topographic Index, Mean Annual Precipitation, Mean Annual Temperature, Solar Radiation, Mean Annual Water Deficit, Population Density, Income, Proximity to Roads, Continuous Vegetation Field, and Land Use and Land Cover. The RF's performance was compared to previously validated results for Classification and Regression Trees (CART) using the coefficient of determination (R^2) and the root mean square error (RMSE). The RF exhibited slightly superior results in both metrics, with an R^2 of 0.9718 compared to CART's 0.88, and an RMSE of 0.0162 against 0.55, suggesting a greater ability to capture complexities in the data and generate more accurate predictions.*

Keywords: *Remote Sensing, Machine Learning, Random Forest, Forest Fire Prediction*