



**UNIVERSIDADE FEDERAL DO OESTE DO PARÁ  
PRÓ-REITORIA DE PESQUISA, PÓS-GRADUAÇÃO E INOVAÇÃO  
TECNOLÓGICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM RECURSOS NATURAIS DA AMAZÔNIA**

**KEMUEL MACIEL FREITAS**

**PREDIÇÃO DA SUSCEPTIBILIDADE A INCÊNDIOS NA APA TRIUNFO DO  
XINGU COM FERRAMENTAS DE MACHINE LEARNING**

**SANTARÉM – PA  
2024**

**KEMUEL MACIEL FREITAS**

**PREDIÇÃO DA SUSCEPTIBILIDADE A INCÊNDIOS NA APA TRIUNFO DO  
XINGU COM FERRAMENTAS DE MACHINE LEARNING**

Dissertação apresentada ao Programa de Pós-Graduação em Recursos Naturais da Amazônia da Universidade Federal do Oeste do Pará, como requisito para obtenção do título de mestre.

Orientador: Prof. Dr. Anderson Alvarenga de Moura  
Meneses

Coorientador: Prof. Dr. Ronie Silva Juvanhol

**SANTARÉM – PA  
2024**

**Dados Internacionais de Catalogação-na-Publicação (CIP)**  
**Sistema Integrado de Bibliotecas – SIBI/UFOPA**

---

F866p Freitas, Kemuel Maciel  
Predição da susceptibilidade a incêndios na Apa Triunfo do Xingu co ferramentas de machine learning./ Kemuel Maciel Freitas. - Santarém, 2024.  
113 p. : il.  
Inclui bibliografias.

Orientador: Anderson Alvarenga de Moura Meneses.

Coorientador: Ronie Silva Juvanhol.

Dissertação (Mestrado) – Universidade Federal do Oeste do Pará, Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação tecnológica, Programa de Pós-Graduação em Recursos Naturais da Amazônia.

1. Sensoriamento Remoto. 2. Predição de Incêndios. 3. Amazônia. 4. Random Forest. 5. XGBoost. I. Meneses, Anderson Alvarenga de Moura, *orient.* II. Juvanhol, Ronie Silva, *coorient.* III. Título.

CDD: 23 ed. 621.3678

---

Bibliotecária - Documentalista: Cátia Alvarez – CRB/2 843



*Universidade Federal do Oeste do Pará*  
**PROGRAMA DE PÓS-GRADUAÇÃO EM RECURSOS NATURAIS DA  
AMAZÔNIA**

**ATA Nº 198**

Aos 19 (dezenove) dias do mês de agosto de 2024 (dois mil e vinte e quatro), às 14h30, no Auditório NTB / Campus Tapajós, realizou-se a sessão de defesa pública da dissertação ao grau de mestre em Ciências Ambientais - área de concentração Estudos e Manejos de Ecossistemas Amazônicos, de autoria de KEMUEL MACIEL FREITAS, intitulada PREDIÇÃO DA SUSCEPTIBILIDADE A INCÊNDIOS NA APA TRIUNFO DO XINGU COM FERRAMENTAS DE MACHINE LEARNING. A Banca Examinadora constituiu-se por: RAPHAEL PABLO TAPAJOS SILVA (UFOPA - Examinador Interno); MILENA MARILIA NOGUEIRA DE ANDRADE (UFRA - Examinadora Externa à Instituição); MARLA TERESINHA BARBOSA GELLER (ULBRA - Examinadora Externa à Instituição). Seguindo o regimento interno do curso, o presidente da banca Dr. Anderson Alvarenga de Moura Meneses (Orientador) abriu a sessão, passando a palavra para o mestrando, que fez a exposição do seu trabalho. Findada a arguição, a banca examinadora se reuniu, sem a presença do candidato e do público, tendo deliberado por sua: (X) APROVAÇÃO, condicionado às correções sugeridas pela Banca Examinadora e entrega da versão corrigida no prazo de 30 dias; ( ) REPROVAÇÃO, condicionado as correções sugeridas pela Banca Examinadora e reapresentação da dissertação no prazo de 60 dias. Nada mais havendo por constar, lavrou-se e fez-se a leitura da presente ata que segue assinada pelos membros da Banca Examinadora, Presidente (orientador) e mestrando. Santarém-PA. 19 de agosto de 2024.

Documento assinado digitalmente  
**gov.br** MILENA MARILIA NOGUEIRA DE ANDRADE  
Data: 21/08/2024 10:11:49-0300  
Verifique em <https://validar.it.gov.br>

**Dra. MILENA MARILIA NOGUEIRA DE ANDRADE, UFRA**

Examinadora Externa à Instituição

Documento assinado digitalmente  
**gov.br** MARLA TERESINHA BARBOSA GELLER  
Data: 22/08/2024 16:49:13-0300  
Verifique em <https://validar.it.gov.br>

**Dra. MARLA TERESINHA BARBOSA GELLER, ULBRA**

Examinadora Externa à Instituição

Documento assinado digitalmente  
**gov.br** RAPHAEL PABLO TAPAJOS SILVA  
Data: 23/08/2024 09:32:01-0300  
Verifique em <https://validar.it.gov.br>

**Dr. RAPHAEL PABLO TAPAJOS SILVA, UFOPA**

Examinador Interno

Documento assinado digitalmente  
**gov.br** ANDERSON ALVARENGA DE MOURA MENESES  
Data: 21/08/2024 09:19:01-0300  
Verifique em <https://validar.it.gov.br>

**Dr. ANDERSON ALVARENGA DE MOURA MENESES, UFOPA**

Presidente

**KEMUEL MACIEL FREITAS**

Mestrando

Documento assinado digitalmente  
**gov.br** KEMUEL MACIEL FREITAS  
Data: 23/08/2024 11:30:55-0300  
Verifique em <https://validar.it.gov.br>

## AGRADECIMENTO

Primeiramente, a Deus pela oportunidade de chegar aonde cheguei, pelas portas que me foram abertas e pelas que me foram fechadas. Acredito que o Senhor me conduzirá para um lugar onde serei abençoado, e que cada "não" significa que Tu tens algo melhor. Continuarei confiando que se o mundo em mim desmoronar e nada eu puder mudar, poderei descansar pois sei que o todo o mais fará;

À minha família, meus pais e meu irmão que estiveram ao meu lado nessa caminhada e nunca deixaram de apoiar o que eu tinha proposto no meu coração, sempre dedicando tempo, atenção e esforço para que eu pudesse seguir em frente;

Aos meus orientadores Prof. Dr. Anderson Meneses e Prof. Dr. Ronie Juvanhol pela parceria, dedicação e aprendizados durante o período do mestrado, e também pela paciência e ensinamentos nos mais diversos assuntos;

Aos meus amigos de mestrado; colegas e amigos de laboratório; e alunos do estágio docência, por terem tornado essa jornada mais leve, por todas as trocas de conhecimento, brincadeiras e risadas, lanches e conversas;

Aos meus amigos e irmãos de Porto Velho, que foram um presente de Deus para uma mente cansada e um coração abatido, e que trouxeram alegria e renovo para o final dessa jornada;

A todos aqueles que fizeram parte dessa caminhada acadêmica que já dura mais de 10 anos, professores, amigos e colegas da UECE, UFRA E UFOPA, que tiveram contribuição importante em todo esse processo, e que continuarão sendo parte importante da minha formação pessoal e profissional.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001

*“Três Coisas para nunca esquecer:  
- Você é incrível, mesmo que não acredite  
- Você nunca estará só  
- Deus tem um propósito para todas as coisas  
debaixo do sol”*

(Autor Desconhecido)

*“Por que Deus criou esse grande universo?  
- Para o Louvor da Sua Glória!  
Por que Deus criou o homem, afinal de contas?  
- Para o Louvor da Sua Glória!  
Por que Deus fez todas as coisas?  
- Para o Louvor da Sua Glória!”*

(Adauto J. B. Lourenço)

## RESUMO

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, ferramentas de machine learning têm se mostrado promissoras no contexto da previsão de incêndios, de modo que diversos modelos têm sido concebidos nos mais diversos contextos e locais, com aplicações também adaptadas ao contexto brasileiro. Esta pesquisa objetiva realizar o mapeamento de áreas de susceptibilidade a incêndio florestal na Área de Proteção Ambiental Triunfo do Xingu-PA, utilizando algoritmos de aprendizado de máquina para verificar a influência de fatores ambientais, topográficos e socioeconômicos na ocorrência das queimadas. Para isso, foram utilizados os modelos de regressão Random Forest e Extreme Gradient Boosting para prever os valores de densidade de kernel calculada sobre 15.291 pontos de queima confirmados entre 2010 e 2020, utilizando 11 fatores preditores, entre eles, Elevação, Declividade, Orientação de Vertentes, Topographical Wetness Index, Precipitação, Temperatura, Proximidade de Centros Urbanos, Proximidade de Estradas, Uso e Cobertura da terra, Campo Contínuo de Vegetação e o Normalized Difference Vegetation Index. Para avaliar o desempenho dos algoritmos foram utilizadas as métricas de Erro Absoluto Médio (MAE), Raiz do Erro Quadrático Médio (RMSE) e o Coeficiente de Determinação ( $R^2$ ). Os resultados de teste mostraram que os modelos tiveram desempenhos similares, e que tanto o Random Forest (RMSE = 36,26, MAE = 17,45 e  $R^2 = 0,99$ ) quanto o Extreme Gradient Boosting (RMSE = 35,73, MAE = 18,74 e  $R^2 = 0,99$ ) apresentaram uma boa capacidade de previsão, com diferença estatisticamente significativa apenas nos valores de MAE, segundo o teste estatístico de Mann-Whitney. O mapa elaborado apresentou áreas de susceptibilidade alta e muito alta ocupando 39% da área total da unidade de conservação, localizadas principalmente nas regiões central leste e central oeste. As variáveis com maior importância e contribuição para o modelo final foram variáveis ambientais e socioeconômicas, com destaque para a precipitação, a distância de áreas habitadas e o tipo de uso da terra.

**Palavras-Chave:** Sensoriamento Remoto. Previsão de Incêndios. Amazônia. Random Forest. XGBoost.

## ABSTRACT

The behavior and characteristics of fire are determined by a variety of factors, including topography, vegetation, climate, and socioeconomic aspects, which makes it a complex task to predict in which areas resources should be allocated for combatting and preventing forest fires. Machine learning tools have demonstrated promising results for fire prediction, which have included the generation of models that have been developed across a large range of contexts and locations, with applications to the Brazilian context. This research aims to map areas susceptible to forest fires within the Triunfo do Xingu Environmental Protection Area, employing machine learning algorithms to ascertain the influence of environmental, topographic and socioeconomic factors on fire occurrence. For this purpose, the Random Forest and Extreme Gradient Boosting regression models were used to predict kernel density values calculated over 15,291 confirmed burn points between 2010 and 2020, using 11 predictor factors, including Elevation, Slope, Aspect, Topographic Wetness Index, Precipitation, Temperature, Proximity to Inhabited areas, Proximity to Roads, Land Use and Cover, Vegetation Continuous Fields, and the Normalized Difference Vegetation Index. To evaluate the performance of the algorithms, the metrics used were Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the Coefficient of Determination ( $R^2$ ). The test results showed that the models had similar performance, and both the Random Forest (RMSE = 36.26, MAE = 17.45, and  $R^2 = 0.99$ ) and the Extreme Gradient Boosting (RMSE = 35.73, MAE = 18.74, and  $R^2 = 0.99$ ) demonstrated good predictive capacity, with a statistically significant difference only in the MAE values, according to the Mann-Whitney statistical test. The elaborated map presented areas of high and very high susceptibility occupying 39% of the total area of the conservation unit, mainly located in the central-east and central-west regions. The variables with the greatest importance and contribution to the final model were environmental and socioeconomic variables, notably precipitation, distance from inhabited areas, and land use type.

**Keywords:** Remote Sensing. Forest Fire Prediction. Amazon. Random Forest. XGBoost.

## LISTA DE FIGURAS

<b>Figura 1:</b> Limites Geográficos da Amazônia Legal .....	27
<b>Figura 2:</b> Total de Focos de Queimadas Ativos observados na Região da Amazônia Legal no período de 2005 a 2020.....	28
<b>Figura 3:</b> Quantidade de Focos de Calor nas cinco Unidades de Conservação da Amazônia Legal mais atingidas no período de 2019-2021 .....	28
<b>Figura 4:</b> Áreas do conhecimento nas quais podem ser aplicados os conceitos da Inteligência Artificial.....	31
<b>Figura 5:</b> Diferenças entre Inteligência Artificial, Aprendizagem de Máquina e Aprendizagem Profunda.....	32
<b>Figura 6:</b> Relação entre uma Rede Neural Biológica e uma Rede Neural Artificial ..	32
<b>Figura 7:</b> Funcionamento geral de algoritmos baseados em árvores.....	33
<b>Figura 8:</b> Funcionamento de Algoritmos Ensemble baseados em Bagging e em Boosting .....	35
<b>Figura 9:</b> Funcionamento do Processo de Decisão entre as diferentes árvores do Random Forest.....	36
<b>Figura 10:</b> Funcionamento do Processo de Decisão entre as diferentes árvores do XGBoost.....	37
<b>Figura 11:</b> Representação Visual dos Principais Procedimentos e Ferramentas Envolvidos .....	43
<b>Figura 12:</b> Unidades de Conservação Federais, Estaduais e Territórios Indígenas que compõem o Mosaico de Áreas Protegidas Terra do Meio.....	44
<b>Figura 13:</b> Localização Geográfica da Área de Estudo, a APA Trinfo do Xingu, no Estado do Pará, Brasil.....	45
<b>Figura 14:</b> Caracterização dos Solos e da Vegetação da Área de Proteção Ambiental Triunfo do Xingu, no Estado do Pará, Brasil.....	46
<b>Figura 15:</b> Caracterização Geomorfológica da Área de Proteção Ambiental Triunfo do Xingu, no Estado do Pará, Brasil.....	46

<b>Figura 16:</b> Mapa de Curvas de Nível da Área de Proteção Ambiental Triunfo do Xingu, no Estado do Pará, Brasil.....	47
<b>Figura 17:</b> Quantidade de focos de queimada registrados pelo Instituto Nacional de Pesquisas Espaciais (INPE) confirmados na APA Triunfo do Xingu entre os anos de 2010 e 2020 .....	48
<b>Figura 18:</b> Características de (A) Elevação, (B) Declividade, (C) Orientação de Vertentes e (D) Índice Topográfico Composto, para a área da APA Triunfo do Xingu .....	52
<b>Figura 19:</b> Comunidades e Estradas Localizadas na APA Triunfo do Xingu e no seu entorno .....	53
<b>Figura 20:</b> Características de (A) Uso e Cobertura da Terra, (B) Índice de Vegetação por Diferença Normalizada e (C) Campo Contínuo de Vegetação, da APA Triunfo do Xingu .....	55
<b>Figura 21:</b> Características de (A) Temperatura de Superfície e (B) Precipitação Média Anual, da APA Triunfo do Xingu.....	56
<b>Figura 22:</b> Divisão dos Dados de Entrada em conjuntos de Treino/Teste para realização da validação cruzada e Conjunto de Validação para testar a capacidade de generalização do modelo .....	58
<b>Figura 23:</b> Matriz de Correlação Entre as variáveis consideradas para serem preditoras de Incêndios Florestais.....	67
<b>Figura 24:</b> Boxplots dos valores de validação cruzada para RMSE, MAE e $R^2$ para os algoritmos RF e XGB usando os dados de teste/treino.....	70
<b>Figura 25:</b> Boxplots dos valores de validação cruzada para RMSE, MAE e $R^2$ para o algoritmo XGB usando os dados de validação.....	72
<b>Figura 26:</b> Mapa de Predição da Susceptibilidade a Incêndios Florestais na APA Triunfo do Xingu, Pará, Brasil.....	73
<b>Figura 27:</b> Contribuição Média de cada variável preditora para as predições realizadas pelo modelo XGB de Predição de Susceptibilidade a Incêndios Florestais .....	78

## LISTA DE QUADROS

<b>Quadro 1:</b> Relação de variáveis topográficas, climáticas, ambientais e antrópicas com a ocorrência e a propagação de Incêndios Florestais .....	23
<b>Quadro 2:</b> Impactos ocasionados por incêndios florestais .....	26
<b>Quadro 3:</b> Exemplos de Trabalhos de Predição de Incêndios que utilizam de metodologias de IA, ML e DL. ....	39
<b>Quadro 4:</b> Variáveis consideradas como preditoras na análise dos modelos de Predição de Incêndios.....	50
<b>Quadro 5:</b> Variações de parâmetros testadas para os modelos Random Forest e XGBoost utilizando GridSearchCV.....	59
<b>Quadro 6:</b> Comparação em termos percentuais da Importância de cada variável para os Modelos RF e CART .....	66
<b>Quadro 7:</b> Melhores Combinações de parâmetros encontradas para os modelos RF e XGB utilizando GridSeachCV .....	68
<b>Quadro 8:</b> Caracterização das Áreas de Muito Baixa e Baixa Susceptibilidade em relação às Áreas de Muito Alta Susceptibilidade a incêndios florestais considerando as variáveis preditoras estudadas .....	75
<b>Quadro 9:</b> Importância Percentual de cada variável preditora para o modelo de predição XGBoost .....	77

## LISTA DE TABELAS

<b>Tabela 1:</b> Comparação dos resultados obtidos a partir da validação cruzada ( $k = 10$ ) para o algoritmo RF com o resultado médio obtido por Juvanhol et al. (2023) para o Algoritmo CART utilizando o mesmo conjunto de dados .....	65
<b>Tabela 2:</b> Valores da Validação Cruzada, Média, Desvio Padrão, Mediana, Mínimos e Máximos de Raiz do RMSE, MAE e $R^2$ para os algoritmos RF e XGB usando os dados de teste/treino.....	69
<b>Tabela 3:</b> Valores da Validação Cruzada, Média, Desvio Padrão, Mediana, Mínimos e Máximos de RMSE, MAE e $R^2$ para o algoritmo XGB usando os dados de validação .....	71

## LISTA DE SIGLAS

APA	Área De Proteção Ambiental
BDQUEIMADAS	Banco de Dados de Queimadas
CART	<i>Classification and Regression Tree</i>
CHIRPS	<i>Climate Hazards Group Infrared Precipitation with Station Data</i>
CO <sub>2</sub>	Dióxido de Carbono
DL	<i>Deep Learning</i>
DNIT	Departamento Nacional de Infraestrutura e Transportes
ESEC	Estação Ecológica
EVI	<i>Enhanced Vegetation Index</i>
FLONA	Floresta Nacional
FLOTA	Floresta Estadual
FPAR	<i>Fraction of Photosynthetically Active Radiation</i>
GBDT	<i>Gradient Boosted Decision Tree</i>
GEE	<i>Google Earth Engine</i>
IA	Inteligência Artificial
IBAMA	Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis
IBGE	Instituto Brasileiro de Geografia e Estatística
IMAZON	Instituto do Homem e Meio Ambiente da Amazônia
INPE	Instituto Nacional de Pesquisas Espaciais
IPAM	Instituto de Pesquisa Ambiental da Amazônia
LAI	<i>Leaf Area Index</i>
LST	<i>Land Surface Temperature</i>
LST DAY	<i>Land Surface Temperature during the Day</i>
LSTM	<i>Long-Short Term Memory</i>
LST NIGHT	<i>Land Surface Temperature during the Night</i>
MAE	<i>Mean Absolute Error</i>
MDE	Modelo Digital de Elevação
MG	Minas Gerais
ML	<i>Machine Learning</i>
MODIS	<i>Moderate-Resolution Imaging Spectroradiometer</i>
MRIF	Mapa de Risco de Incêndio Florestal

MSE	<i>Mean Squared Error</i>
NDMI	<i>Normalized Difference Moisture Index</i>
NDVI	<i>Normalized Difference Vegetation Index</i>
NDWI	<i>Normalized Difference Water Index</i>
NIR	<i>Near Infrared</i>
PA	Pará
PARNA	Parque Nacional
PREVFOGO	Sistema Nacional e Prevenção e Combate ao Fogo
PROARCO	Programa de Controle e Prevenção do Fogo para o Arco do Desmatamento
R <sup>2</sup>	Coeficiente de Determinação
RESEX	Reserva Extrativista
RF	<i>Random Forest</i>
RMSE	<i>Root Mean Squared Error</i>
RN	<i>Redes Neurais</i>
SHAP	<i>Shapley Additive Explanations</i>
SIG	Sistema de Informações Geográficas
SNUC	Sistema Nacional de Unidades de Conservação
SR	Sensoriamento Remoto
SRTM	<i>Shuttle Radar Topographic Mission</i>
SVM	<i>Support Vector Machine</i>
TI	Território Indígena
TWI	<i>Topographic Wetness Index</i>
UC	Unidade de Conservação
USGS	<i>United States Geological Survey</i>
VCF	<i>Vegetation Continuous Fields</i>
XGB	<i>Extreme Gradient Boosting</i>

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	17
<b>2 OBJETIVOS</b> .....	19
<b>2.1 Objetivo Geral</b> .....	19
<b>2.2 Objetivos Específicos</b> .....	19
<b>3 REFERENCIAL TEÓRICO</b> .....	20
<b>3.1 Incêndios e Queimadas</b> .....	20
<b>3.2 Características e Comportamento do Fogo</b> .....	22
<b>3.3 Impactos causados por Incêndios Florestais</b> .....	24
<b>3.4 Ocorrência de Incêndios na Amazônia</b> .....	26
<b>3.5 Mapeamento de Incêndios Florestais</b> .....	29
<b>3.6 Inteligência Artificial</b> .....	30
3.6.1 Árvores de Classificação e Regressão (CART) .....	33
3.6.2 Random Forest (RF) .....	34
3.6.3 Extreme Gradient Boosting (XGB) .....	36
<b>3.7 Machine Learning na Predição de Incêndios</b> .....	37
<b>4 MATERIAL E MÉTODOS</b> .....	42
<b>4.1 Área de Estudo</b> .....	44
<b>4.2 Densidade (Kernel) de Queimadas – Variável Resposta</b> .....	47
<b>4.3 Variáveis Predictoras</b> .....	49
4.3.1 Variáveis Topográficas .....	51
4.3.2 Variáveis Socioeconômicas .....	52
4.3.3 Variáveis Ambientais .....	53
<b>4.4 Seleção de Variáveis</b> .....	56
<b>4.5 Processamento dos Dados</b> .....	57
4.5.1 Pré-Processamento .....	57
4.5.2 Validação Cruzada – Dados de Treino e de Validação .....	57
<b>4.6 Configurações</b> .....	58
4.6.1 Configuração dos Modelos Machine Learning .....	58
4.6.2 Configuração de Hardware e Software .....	59
<b>4.7 Avaliação de Desempenho</b> .....	60
4.7.1 Métricas de Desempenho .....	60
4.7.2 Análises Estatísticas .....	61
<b>4.8 Testes Preliminares – Benchmark</b> .....	62
<b>4.9 Predição da Susceptibilidade a Incêndios</b> .....	63

4.9.1 Mapa de Predição da Susceptibilidade de Incêndios Florestais .....	63
4.9.2 Importância e Contribuição das Variáveis.....	63
<b>5 RESULTADOS E DISCUSSÃO .....</b>	<b>64</b>
<b>5.1 Testes Preliminares – Benchmark.....</b>	<b>64</b>
<b>5.2 Estudo de Caso – APA Triunfo do Xingu.....</b>	<b>66</b>
5.2.1 Seleção de Variáveis .....	66
5.2.2 Parâmetros dos Modelos Machine Learning.....	67
5.2.3 Desempenho dos Algoritmos.....	68
5.2.4 Resultados dos Dados de Validação .....	71
5.2.5 Mapa de Predição da Susceptibilidade de Incêndios Florestais .....	72
5.2.6 Influência e Contribuição das Variáveis .....	76
<b>6 CONSIDERAÇÕES FINAIS .....</b>	<b>80</b>
<b>REFERÊNCIAS.....</b>	<b>81</b>
<b>APÊNDICE A – ARTIGO APRESENTADO NO XXVI ENCONTRO NACIONAL DE MODELAGEM COMPUTACIONAL.....</b>	<b>98</b>
<b>APÊNDICE B – RESULTADOS DO GRIDSEARCHCV PARA OS ALGORITMOS RANDOM FOREST E XGBOOST .....</b>	<b>99</b>

## 1 INTRODUÇÃO

A Amazônia é uma das regiões mais importantes do planeta, abrigando cerca de 13% de todas as árvores do mundo (Gomes *et al.*, 2019), estocando entre 40 e 50% do carbono terrestre total, sendo uma gigantesca reserva de carbono (Ferreira *et al.*, 2018), e também um dos principais fatores reguladores do ciclo da água e do clima, tanto em escala regional quanto global (Ellwanger *et al.*, 2020), além de abrigar uma enorme diversidade de recursos naturais e culturais (Facchinelli *et al.*, 2019).

Ao longo da história, as populações que habitam a Amazônia têm utilizado o fogo como uma ferramenta para realizar o manejo do solo, limpeza de plantações, manutenção de pastagens, etc. (Sá *et al.*, 2007). Esse uso do fogo de forma controlada, bem como a ocorrência natural do fogo em alguns ambientes e situações, tem suscitado inúmeros benefícios para as atividades humanas, e também em determinados ecossistemas que já estão acostumados à ocorrência natural do fogo, como no cerrado, onde as plantas possuem adaptações morfológicas e/ou fisiológicas à essa “passagem” do fogo (Simon *et al.*, 2009).

Contudo, com a crescente pressão da agropecuária e do desmatamento na região, o fogo muitas vezes foge de controle, transformando-se em incêndios florestais destrutivos, que podem se propagar rapidamente e causar danos irreparáveis à floresta, ao meio ambiente e à sociedade ao redor (Fonseca-Morello *et al.*, 2017). Ao perceber o quanto a Amazônia tem sido impactada pelas queimadas e, conseqüentemente, pelo desmatamento, é importante que sejam adotadas estratégias que possam proteger os recursos ambientais, bem como as funções ecológicas exercidas pela floresta. 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 gerenciamento de uma área tão extensa como a Amazônia Legal, com aproximadamente 5 milhões de km<sup>2</sup>, correspondendo a 58,9% do território nacional (IBGE, 2023), precisa otimizar o uso dos recursos disponíveis, particularmente, tempo, mão-de-obra e o dinheiro, para reduzir a vulnerabilidade e melhorar o processo de tomada de decisão em relação à prevenção, monitoramento e combate dos riscos ambientais (Almeida *et al.*, 2020; Ozenen Kavlak *et al.*, 2021).

O monitoramento ativo do fogo na Amazônia foi iniciado por volta dos anos 1980 por meio do Instituto Nacional de Pesquisas Espaciais (INPE). Em 1989, o Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis (IBAMA) estabeleceu o Sistema Nacional de Prevenção e Combate ao Fogo (PREVFOGO) e em maio de 1998, com suporte financeiro do Banco Mundial, foi estabelecido o Programa de Controle e Prevenção do Fogo para o Arco do Desmatamento (PROARCO) para apoiar o planejamento estratégico regional e auxiliar nas atividades de campo do PREVFOGO.

Todavia, novas tecnologias têm se mostrado promissoras na busca por soluções para esses problemas. Neste contexto, o uso de ferramentas de machine learning (ML) tem se destacado para a predição e monitoramento de incêndios florestais (Abid, 2021). A capacidade de analisar grandes volumes de dados em tempo real possibilita a detecção precoce de áreas mais susceptíveis ao fogo, permitindo uma resposta mais rápida e eficiente no combate aos incêndios florestais.

Diversos modelos que utilizam métodos de ML e *deep learning* (DL) para predição de incêndios florestais já têm sido desenvolvidos em diferentes contextos e situações, utilizando Redes Neurais (RN), Algoritmos de Árvores de Regressão e Classificação (*Classification and Regression Tree - CART*), Random Forest (RF), Máquina de Vetores de Suporte (*Support-Vector Machine - SVM*), Redes Neurais Profundas (RNP), Extreme Gradient Boosting (XGB), além de vários outros, na Ásia (Pourghasemi *et al.*, 2020a; 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, ainda há uma carência de que mais estudos sejam desenvolvidos na Amazônia, principalmente em áreas protegidas que já estão sob pressão de desmatamentos e de incêndios já recorrentes, como é o caso da Área de Proteção Ambiental (APA) Triunfo do Xingu, no estado do Pará, apontada pelo INPE por meio do Projeto de Monitoramento do Desmatamento na Amazônia Legal por Satélite – PRODES, como a unidade de conservação mais desmatada na Amazônia Legal e a que mais apresenta focos de queimadas (INPE, 2023).

A partir disso, busca-se por meio desta pesquisa realizar a predição e o mapeamento de áreas de susceptibilidade a incêndio florestal na APA Triunfo do

Xingu-PA, utilizando ferramentas de *machine learning* para verificar a influência de fatores ambientais, topográficos e socioeconômicos na ocorrência das queimadas.

Para isso, esta dissertação está organizado em 6 seções: Introdução – uma abordagem geral das problemáticas envolvidas, bem como das ferramentas e técnicas estudadas e aplicadas; Objetivos – apresentação dos objetivos geral e específicos, gerados a partir das perguntas e hipóteses do trabalho; Referencial Teórico – apresentando alguns conceitos teóricos básicos para a fundamentação deste trabalho, como Incêndios e Queimadas, Características e Comportamento do Fogo, os Impactos causados por incêndios florestais, a ocorrência de incêndios na Amazônia, a Inteligência Artificial e a utilização do ML na Previsão de Incêndios; Material e Métodos – detalhando o fluxograma da pesquisa e os procedimentos empenhados na aquisição dos dados, no seu pré-processamento e tratamento, bem como a seleção e configuração dos modelos utilizados, as métricas utilizadas para avaliar os modelos e os testes estatísticos utilizados, e também procedimento de realização dos testes preliminares; Resultados e Discussão – mostrando os resultados obtidos desde a etapa de testes preliminares, a seleção de variáveis, o desempenho dos modelos nas etapas de teste e de validação, bem como o mapa de predição e a influência e contribuição de cada variável preditora; e por fim, Considerações Finais – com as conclusões obtidas a partir do desenvolvimento da pesquisa até o presente momento.

## **2 OBJETIVOS**

### **2.1 Objetivo Geral**

Realizar a predição e o mapeamento de áreas de susceptibilidade a incêndio florestal na APA Triunfo do Xingu-PA, utilizando algoritmos de aprendizagem de máquina para verificar a influência de fatores ambientais, topográficos e socioeconômicos na ocorrência das queimadas.

### **2.2 Objetivos Específicos**

- Comparar diferentes modelos de aprendizagem de máquina, a partir de diferentes métricas para as análises multicritérios dos fatores envolvidos;
- Identificar o modelo mais adequado para a predição dos pontos com maior ocorrência de focos de incêndio;

- Mapear, em ambiente SIG as áreas de maior susceptibilidade de ocorrência de incêndios florestais, caracterizar as áreas em função dos fatores preditores e analisar a influência desses fatores

### **3 REFERENCIAL TEÓRICO**

Esta seção apresenta uma visão geral sobre diferentes conceitos essenciais para esta dissertação, abordando os conceitos de Queimadas e Incêndios Florestais, as características e o comportamento do fogo, os principais impactos causados pelos incêndios florestais na esfera ambiental, econômica e social, bem como a ocorrência de incêndios na Amazônia e o início do processo de mapeamento de Incêndios Florestais. Também são introduzidos alguns conceitos de Inteligência Artificial em conjunto com explicações sobre os principais modelos utilizados e, finalmente, a utilização de técnicas de ML na predição de incêndios.

#### **3.1 Incêndios e Queimadas**

No que se refere a eventos de fogo, existem duas terminologias amplamente utilizadas: queimadas e incêndios. Queimadas estão relacionadas às práticas agrícolas e florestais que utilizam o fogo de forma controlada e intencional. Essa é uma prática comum para a limpeza de áreas no Brasil, chegando a ser parte de um processo cultural (Sales *et al.*, 2019), 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 (Carrero *et al.*, 2019; Alves e Modesto Junior, 2020).

Por outro lado, Incêndios Florestais se referem aos casos em que o fogo foge do controle e incide sobre qualquer vegetação, podendo ocorrer em escalas locais e regionais, provocados por causas naturais – podendo ser causados pela ação de raios ou por erupções vulcânicas, ou por conta do homem, por meios acidentais ou meios intencionais, e neste caso, criminosos (Andrade e Ferreira, 2019; Berlinck e Batista, 2020).

Ainda segundo Berlinck e Batista (2020), incêndios florestais ocasionados naturalmente são raros em florestas primárias úmidas e fechadas, como é o caso da

Floresta Amazônica, devido às condições de alta pluviosidade e umidade a tornarem, praticamente, não inflamáveis. Em contrapartida, vários estudos têm observado uma relação direta entre desmatamento e aumento de casos de incêndios florestais (Nepstad *et al.*, 1999; Caúla *et al.*, 2015; Andrade *et al.*, 2019; Fuchs, 2020).

Schroeder *et al.* (2009) propuseram outro tipo de classificação para eventos de queima em florestas, dividindo em quatro grupos com comportamentos e origem distintos. O primeiro grupo diz respeito ao uso do fogo para transformar áreas de floresta em áreas que serão usadas para pastagem ou agricultura. Esse tipo de fogo foi denominado de “Fogo de Desmatamento”, sendo um tipo de fogo intencional, usado como uma solução barata em pequenas propriedades familiares, que possibilita a transformação da matéria orgânica seca em cinzas que servirão de fertilizante.

O segundo grupo também representa um outro tipo de fogo usado de forma intencional, o “fogo de manutenção”, usado principalmente no gerenciamento de pastagens e para limpar os resíduos deixados durante a colheita. Também é usado para controlar o aparecimento de ervas daninhas e outras pestes, bem como restaurar uma parte da fertilidade do solo e rejuvenescer as gramíneas. Contudo, depois de alguns anos, essa fertilidade, o vigor do solo, já começa a decair. Diferente do “fogo de desmatamento”, o “fogo de manutenção” é rápido e fraco em intensidade devido à redução constante da quantidade de biomassa usada como combustível.

O terceiro grupo proposto por Schroeder *et al.* (2009) se difere por dois primeiros por ser causado por mecanismos naturais, sendo denominado “Fogo Natural”. A principal causa desse tipo de fogo são os relâmpagos, contudo também pode ser ocasionado, em certas localidades, por deslizamentos de rochas e erupções vulcânicas. Incêndios naturais não são frequentes na Amazônia por conta dos raios, por via de regra, sempre serem acompanhados de chuva, que extingue a ignição inicial e previne que o fogo se propague.

Por fim, o quarto grupo é responsável pela maior parte da ocorrência de incêndios em áreas de floresta, o “Fogo Acidental” representa aquelas queimas de desmatamento ou de manutenção que fogem de controle. Geralmente esse tipo de fogo afeta áreas que já estavam previamente desmatadas ou afetadas, contudo em períodos mais secos e quentes, pode adentrar em áreas mais preservadas. O comportamento do fogo nestes casos está diretamente relacionado ao grau de perturbação ao qual a área foi submetida, de modo que, em florestas relativamente

preservadas, o fogo tem uma intensidade baixa, se move mais lentamente e tende a não se espalhar por áreas muito grandes. Por outro lado, florestas fragmentadas são mais susceptíveis ao fogo por conta da grande quantidade de biomassa seca disponível.

### **3.2 Características e Comportamento do Fogo**

As características, bem como o comportamento do fogo, são determinados pela correlação entre uma série de fatores, desde a cobertura vegetal, relevo, clima e até atividades antrópicas. O uso do solo e a cobertura vegetal determinam, principalmente, a quantidade e a qualidade da biomassa que vai ser consumida como combustível (Moreira *et al.*, 2020). Entre os aspectos topográficos, tanto a declividade como a orientação das encostas podem atuar como fatores de propagação, enquanto a elevação do terreno por estar relacionada com a umidade, atuando na redução dessa propagação (Moreira *et al.*, 2020; Torres *et al.*, 2017). Com relação à declividade, quanto mais inclinado o terreno estiver, maior será a velocidade de propagação do fogo. Da mesma forma que a orientação das vertentes, ou encostas, sinaliza qual a face que estará mais exposta aos raios solares, e conseqüentemente, uma maior quantidade de energia em forma de calor (Ribeiro *et al.*, 2012). Por fim, a elevação do terreno está relacionada direta e indiretamente a vários outros fatores, como a temperatura do ar, a radiação incidente, que podem afetar, tanto positiva quanto negativamente sobre a ocorrência de incêndios (Messias e Ferreira, 2019).

Variáveis climáticas também podem atuar na propagação no fogo, tais como, a direção do vento e a diminuição da umidade, a temperatura e a precipitação influenciam, principalmente, na umidade daqueles materiais que poderão servir como combustíveis para o fogo (serrapilheira, gramíneas, galhos ou qualquer madeira que estejam no sub-bosque), tornando-os mais ou menos inflamáveis, favorecendo ou não, a ignição e a propagação do fogo (Aragão *et al.*, 2018; Venkatesh *et al.*, 2020). Os fatores relacionais ao clima afetam principalmente a umidade. Da mesma forma, a variação no uso e ocupação do solo também influenciam na cobertura vegetal, que por sua vez pode interferir no comportamento do fogo, tanto como combustível quanto como meio de propagação (Bacani, 2016).

**Quadro 1:** Relação de variáveis topográficas, climáticas, ambientais e antrópicas com a ocorrência e a propagação de Incêndios Florestais

	Variáveis	Relação com o Fogo
Topográficas	Elevação	Relacionada a diferentes fatores, como a umidade, a temperatura, padrões de ventos. Geralmente áreas mais elevadas são mais secas, facilitando a ignição;
	Declividade	Quanto mais inclinado o terreno estiver, maior será a velocidade de propagação do fogo;
	Orientação de Vertentes	No caso da Amazônia, vertentes voltadas para o Norte tendem a ser mais secas e quentes, facilitando a ocorrência de incêndios;
Climáticas	Temperatura	Temperaturas mais elevadas afetam o estado do material combustível, deixando a vegetação mais seca;
	Umidade	Baixos níveis de umidade atmosférica e do solo também influenciam o material combustível, tornando-o mais inflamável;
	Precipitação	Maiores volumes de chuva mantêm a vegetação úmida, diminuindo a ocorrência de incêndios;
Ambientais	Uso e Cobertura do Solo	O tipo de cobertura do solo influencia a cobertura vegetal, de modo que áreas de vegetação seca ou de monoculturas podem favorecer a propagação de incêndios;
	Cobertura Florestal	Áreas de maior cobertura florestal, com vegetação mais densa, tendem a ter uma proteção natural contra esses eventos. Contudo, a recorrência desses eventos diminui essa capacidade de proteção;
	Saúde da Vegetação	Relacionada a outros fatores, como a umidade. Vegetações mais saudáveis, tendem a ser menos susceptíveis ao fogo, por conta do alto teor de umidade;
Antrópicas	Comunidades	Regiões mais próximas a atividades humanas estão sujeitas a ocorrência de incêndios, principalmente por conta de queimadas agrícolas, despejo de lixo e negligência no uso do fogo;
	Estradas	Estradas facilitam o acesso a áreas de floresta, aumentando a ocorrência de incêndios de forma acidental ou intencional;

Fonte: Elaborado pelo Autor (2024)

Por fim, fatores antrópicos estão entre os principais causadores de incêndios, estando relacionados principalmente à proximidade de áreas urbanas e da malha viária (Abedi Gheshlaghi *et al.*, 2020). Tais fatores são levados em consideração para determinar as áreas mais propensas a incêndios, na busca de demonstrar que quanto maior a proximidade de tais atividades, maior o potencial de risco, uma vez que elas facilitam o acesso às áreas mais vulneráveis, além de gerar um maior fluxo de pessoas (Venkatesh *et al.*, 2020; Moreira *et al.*, 2020).

### **3.3 Impactos causados por Incêndios Florestais**

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 e 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).

Olhando para os impactos ambientais, incêndios podem aumentar a vulnerabilidade da floresta à recorrência de incêndios (Gabardo *et al.*, 2021). Podem ainda causar mudanças na biodiversidade, incluindo a mortalidade de árvores, mudanças na composição florestal e na disponibilidade de frutas e impactos na fauna. Além disso, podem ocasionar mudanças na disponibilidade de nutrientes no solo, influenciando na recuperação de áreas e de ocasionarem a perda da qualidade do solo e da água, impactando o ciclo hidrológico (Oliveira *et al.*, 2020; Ladislau *et al.*, 2021).

Os distúrbios causados pelo fogo também acabam reduzindo a abundância das comunidades de invertebrados, diminuindo a disponibilidade de alimento para algumas espécies de aves. Da mesma forma, o calor produzido pode gerar um estresse nas árvores, reduzindo a disponibilidade de alimento para espécies frugívoras.

Incêndios de grandes proporções podem ter consequências climáticas observadas tanto em escala local, regional e também global. Elas podem ser percebidas, principalmente, no aumento da emissão de CO<sub>2</sub>, na medida que o fogo é

responsável por cerca de um terço das emissões de carbono provenientes do desmatamento (Aragão *et al.*, 2018), mas também pelo aumento nas emissões de metano, assim como de outros gases e substâncias, sendo capaz de afetar também outros elementos da superfície, como o albedo, evapotranspiração e o fluxo de calor, além de ser responsável por emitir partículas de aerossóis capazes de alterar a composição da atmosfera amazônica (Santos *et al.*, 2017).

Entre os impactos econômicos diretos associados ao fogo que foge do controle temos a perda de gado e de colheitas (Edgeley e Paveglio, 2017), além do dano às infraestruturas. Numa escala geográfica maior, podem acontecer o fechamento de aeroportos devido à fumaça e à falta de energia (Rodrigues *et al.*, 2020). As queimadas também contribuem para a desvalorização das florestas para sociedade, visto que influenciam o investimento em áreas rurais e também diminuem a produção e aumentam a mortalidade de espécies que são importantes fontes de produtos florestais não-madeireiros (Duff *et al.*, 2019).

Dentre os principais impactos que estão relacionados à esfera social, se destaca o aumento nos casos de doenças respiratórias nas áreas mais próximas, ocasionadas principalmente pelo aumento na emissão de material particulado que podem ser inalados e que ocasionam problemas respiratórios e cardiovasculares (Johnston *et al.* 2014; Reddington *et al.*, 2015). Fernandes e Terceiro (2022) ao estudarem a influência de queimadas na prevalência de doenças respiratórias infantis no hospital de Marabá – PA, apontaram que existe uma correlação significativa entre índices de queimadas e doenças do trato respiratório. Da mesma forma, Castro *et al.* (2009) ao estudarem a tendência da mortalidade por doenças respiratórias em idosos no Estado de Rondônia, também observaram correlação positiva e significativa entre o número de focos de queimadas e as taxas de mortalidade.

Outras situações que também acontecem devido ao acúmulo de fumaça são o fechamento de aeroportos (Pahlow *et al.*, 2005; Albar *et al.*, 2018) e de escolas (Aditama, 2000), com a suspensão das aulas. De acordo com o exposto, o Quadro 1 apresenta os principais impactos ocasionados por incêndios florestais no âmbito ambiental, econômico e social.

**Quadro 2:** Impactos ocasionados por incêndios florestais

<b>IMPACTOS</b>	<b>Ambientais</b>	<ul style="list-style-type: none"> <li>• Vulnerabilidade da Floresta à recorrência de incêndios</li> <li>• Mortalidade de Árvores</li> <li>• Mudanças na Composição Florestal</li> <li>• Disponibilidade de Frutas</li> <li>• Impactos na Fauna</li> <li>• Disponibilidade de Nutrientes no solo</li> </ul>
	<b>Econômicos</b>	<ul style="list-style-type: none"> <li>• Perda de rebanhos e colheitas</li> <li>• Danos às Infraestruturas</li> <li>• Fechamento de Aeroportos</li> <li>• Desvalorização das Florestas</li> </ul>
	<b>Sociais</b>	<ul style="list-style-type: none"> <li>• Doenças Respiratórias</li> <li>• Fechamento de Escolas</li> <li>• Empobrecimento da População</li> <li>• Abandono da Terra</li> </ul>

**Fonte:** Elaborado pelo Autor (2024)

### 3.4 Ocorrência de Incêndios na Amazônia

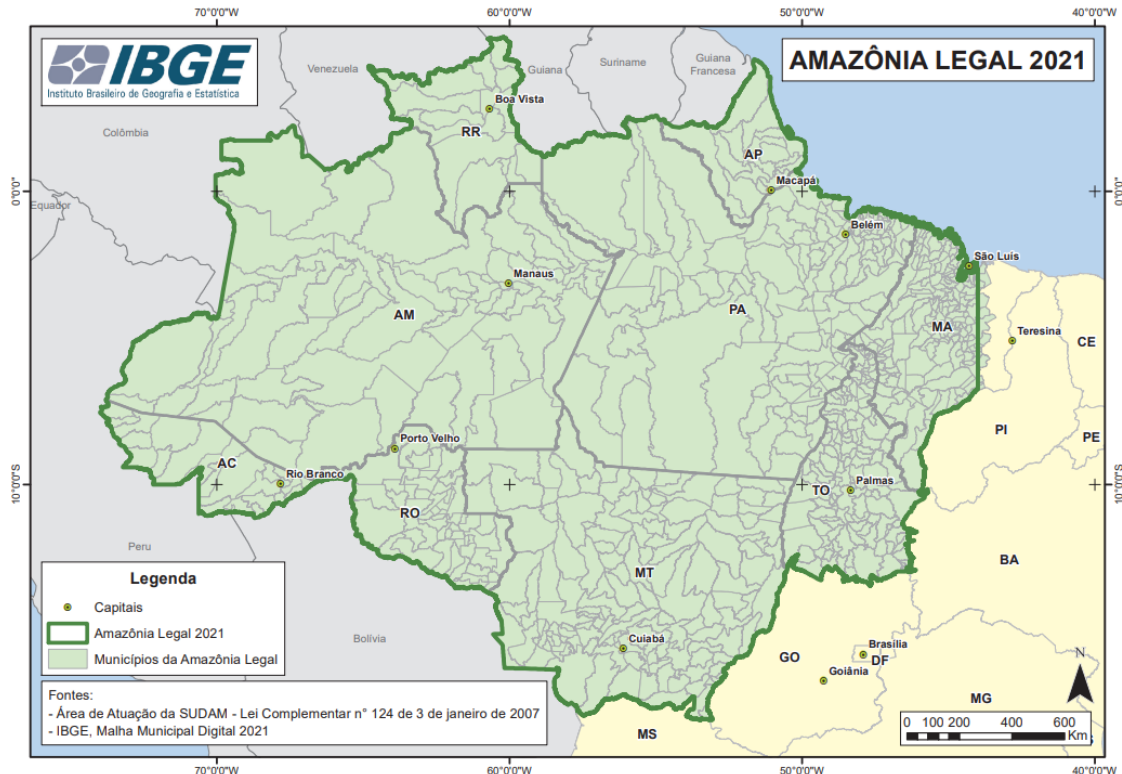
As condições normais da Amazônia impõem uma barreira natural para que os incêndios florestais se desenvolvam. Apesar da temperatura elevada, apresenta altos valores de umidade relativa do ar (Reboita *et al.* 2010) e uma precipitação anual que varia entre 1500 e 2500 mm (Martorano, 1993), além do número consecutivo de dias sem chuva ser baixo, mesmo nos períodos de menor precipitação.

Desse modo, a ocorrência de fogo na Amazônia é marcada pelo contraste entre a baixa ocorrência natural do fogo e uma dominância crescente do fogo relacionado à ocupação humana na região (Latorre *et al.*, 2017). Nas últimas décadas, secas relacionadas à episódios do El Niño combinadas com o número crescente de assentamentos na região e o desenvolvimento da infraestrutura de transporte transformaram o fogo numa grande ameaça ao ecossistema amazônico e ao clima da região (Silva Junior *et al.*, 2019).

Nos últimos 10 anos (2011 – 2021), a Amazônia teve 693.414 km<sup>2</sup> de área queimada, o que traduz a vulnerabilidade deste ecossistema (INPE, 2023). E, manter uma fiscalização constante em uma área tão grande quanto à Amazônia é ineficaz (Figura 1). Contudo, com o desenvolvimento e a recente popularização do Sensoriamento Remoto, bem como dos Sistemas de Informações Geográficas (SIG),

é possível monitorar grandes áreas e coletar, manipular e visualizar um grande volume de dados (Torres *et al.*, 2017).

**Figura 1:** Limites Geográficos da Amazônia Legal



**Fonte:** IBGE (2023)

Ao longo das últimas duas décadas é alarmante a quantidade de focos de queimadas registrados na Amazônia brasileira (Figura 2), especialmente em áreas de conservação. Apesar das secas serem um fator importante na ocorrência de incêndios florestais, principalmente em anos com eventos mais extremos como em 2005, 2007 e 2010 (Alencar *et al.*, 2015; Brando *et al.*, 2019), segundo nota técnica divulgada pelo Instituto de Pesquisa Ambiental da Amazônia – IPAM, grande parte dos focos registrados estão relacionados a áreas recém desmatadas ou já convertidas para uso agropecuário (Alencar *et al.*, 2022).

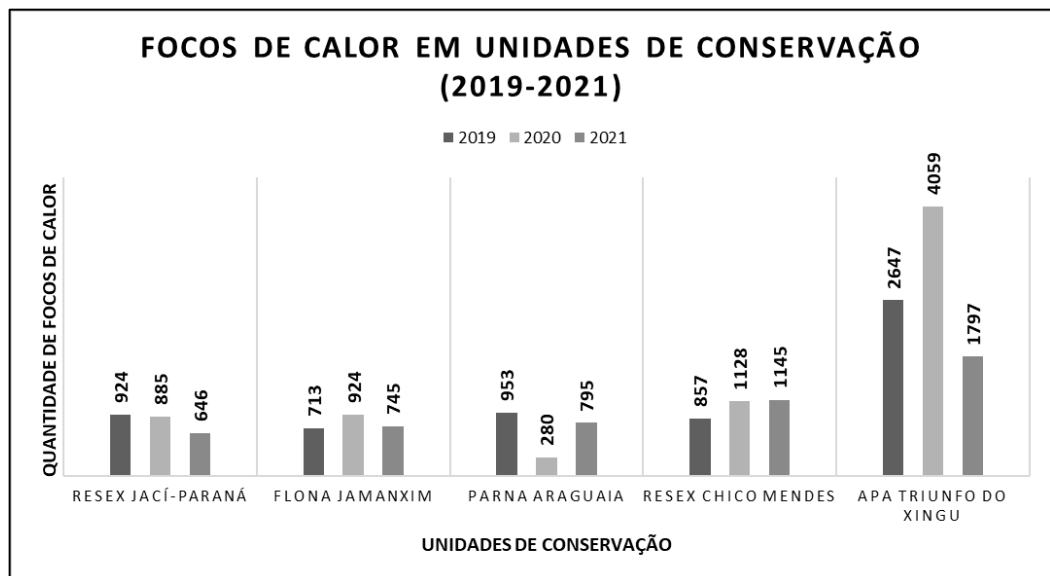
**Figura 2:** Total de Focos de Queimadas Ativos observados na Região da Amazônia Legal no período de 2005 a 2020



Fonte: Programa Queimadas/INPE (2023)

A Amazônia possui 357 Unidades de Conservação (UC), sendo 258 de uso sustentável e 99 de proteção integral, divididas entre as esferas municipal, estadual e federal (MMA, 2022). Dentre todas as UC localizadas na Amazônia, as que estão retratadas na Figura 3 foram as que apresentaram o maior número de focos de calor entre os anos de 2019 e 2021.

**Figura 3:** Quantidade de Focos de Calor nas cinco Unidades de Conservação da Amazônia Legal mais atingidas no período de 2019-2021



Fonte: Synergia Socioambiental (2022)

### 3.5 Mapeamento de Incêndios Florestais

A elaboração de Mapas de Risco de Incêndio Floresta (MRIF), através da utilização de ferramentas de Sistemas de Informações Geográficas (SIG), permite a combinação de vários fatores preditores em um único mapa, que indica as zonas mais ou menos suscetíveis ao fogo dentro de uma área florestal. Dessa forma, os dados e informações resultantes do MRIF tornam-se subsídios para o planejamento e administração dessas áreas.

Além de monitorar é preciso a implementação de ações de prevenção e de detecção do fogo em estágios iniciais. Apesar da influência de vários fatores, é possível modelar a ocorrência e o comportamento do fogo a partir da simulação das interações entre esses fatores e gerar mapas que apontem a susceptibilidade da ocorrência de incêndios florestais em uma determinada área (Mohajane *et al.*, 2021). Várias metodologias têm sido utilizadas para este fim, tanto no Brasil (Torres *et al.*, 2017; Moreira *et al.*, 2020; Almeida *et al.*, 2020), quanto em outros países (Mohajane *et al.*, 2021; Jang *et al.*, 2019; Abdollahi *et al.*, 2018).

Torres *et al.* (2017) compararam quatro modelos de mapeamento de risco de incêndios florestais no município de Viçosa – MG, analisando a influência de fatores preditores como clima, relevo, uso do solo e influência antrópica, atribuindo pesos diferenciados a cada combinação de variáveis, chegando à conclusão de que todos os mapas gerados apresentaram satisfatória eficiência na predição de incêndios.

No entanto, Moreira *et al.* (2020) além de realizarem o mapeamento de áreas suscetíveis a incêndios florestais, apresentaram locais potenciais para a instalação de torres de vigilância desses eventos para a área de dois parques estaduais e seus entornos, no estado de Goiás. O mapa foi elaborado a partir da sobreposição ponderada de fatores relacionados a incêndios florestais, e posteriormente foram determinados locais potenciais para a instalação das torres de observação seguindo determinados fatores: relevo, proximidade de estradas, proximidade de área de risco moderado ou alto e o uso do solo. Os resultados mostraram que mais de 40% da área dos parques possuía alto risco de incêndios e que a instalação de duas torres já possibilitava uma cobertura superior a 70% de toda a área de estudo.

E então, Almeida *et al.* (2020) ao discutirem a inclusão de fatores socioeconômicos na formulação de mapas de risco de incêndio florestal para o

município de Moju – PA, gerando um mapa de risco de desastre por incêndio florestal, notaram que a inclusão ou não de aspectos socioeconômicos gera variações nas áreas de risco quando comparadas entre os dois tipos de mapas.

Considerando estudos realizados em outros países, Jang *et al.* (2019) propuseram um algoritmo baseado na combinação de 3 etapas para a detecção de incêndios florestais na Coreia do Sul, usando os dados do satélite Himawari-8. O algoritmo proposto apresentou uma probabilidade de detecção acima de 93%, sendo capaz de identificar uma quantidade maior de incêndios, enquanto outros algoritmos existentes apresentaram probabilidades entre 50 e 60%, e que não registraram ocorrências de menor escala.

Por fim, Abdollahi *et al.* (2018) sugeriram melhorias para um sistema já existente de previsão de perigo de incêndio florestal baseado exclusivamente em sensoriamento remoto, e sua implementação na província de Alberta, Canadá. O sistema modificado composto pela temperatura da superfície e volume de água precipitados diários, índice de vegetação por diferença normalizada (NDVI - *Normalized Difference Vegetation Index*) e índice de água por diferença normalizada (NDWI - *Normalized Difference Water Index*) compostos de oito dias, foi combinado com uma série temporal de incêndios que ocorreram na região. O resultado da combinação elevou a efetividade na detecção do fogo, com 82% das detecções entre as classes de maior perigo.

### **3.6 Inteligência Artificial**

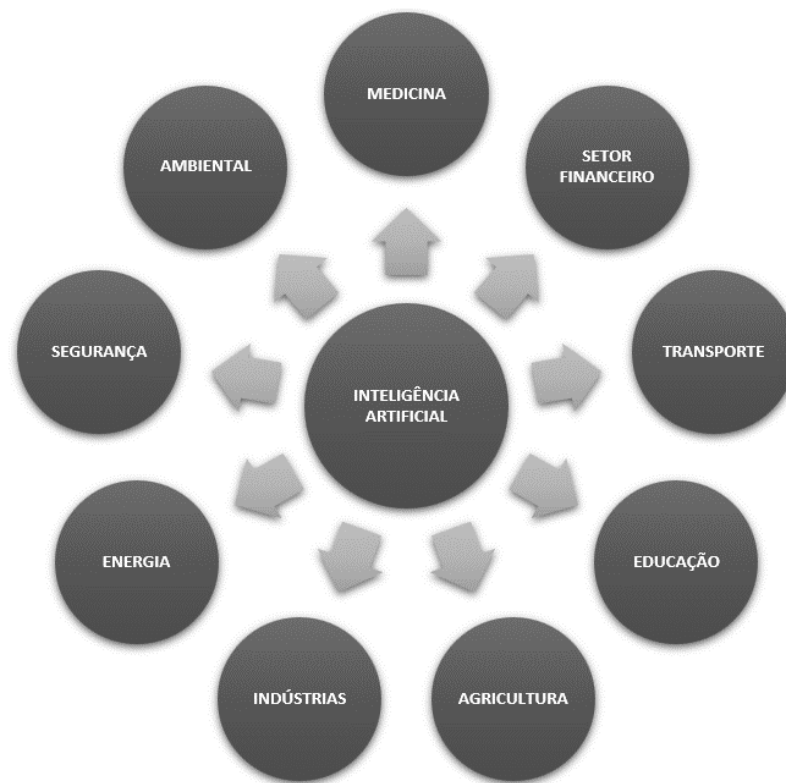
A inteligência artificial surgiu entre as décadas de 1940 e 1950, de modo que a primeira visão geral foi apresentada por Alan Turing. O teste de Turing se baseia no conceito de inteligência, de modo que um computador passará no teste se um interrogador humano não for capaz de distinguir entre as respostas 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).

Assim, 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, 2007). Segundo Haykin (2007), 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.

As aplicações da Inteligência Artificial podem variar entre diversas áreas do conhecimento (Figura 4), podendo estar relacionadas com agricultura (Talaviya *et al.*, 2020; Zhang *et al.*, 2021); energia (Ngarambe *et al.*, 2020; Da Silva *et al.*, 2022); meio ambiente (Aldhyani *et al.*, 2020; Kalantar *et al.*, 2020), entre várias outras.

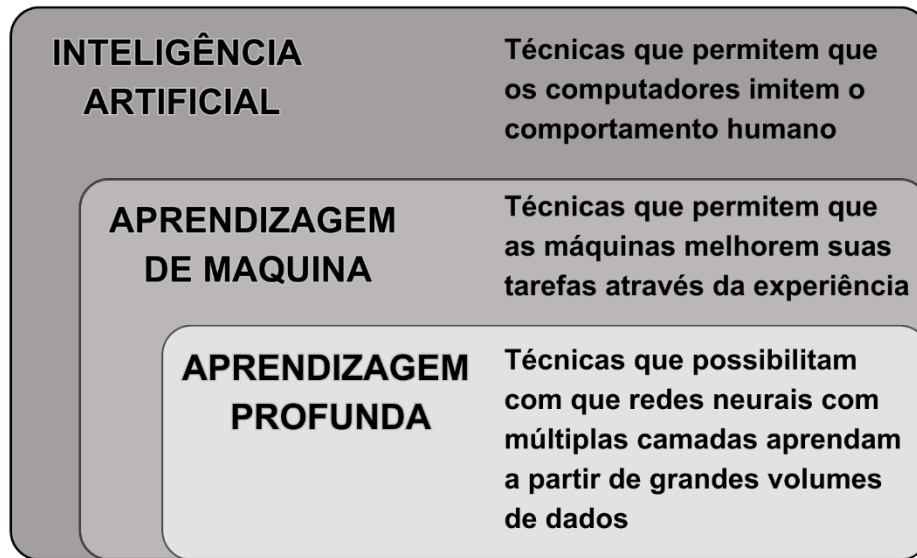
**Figura 4:** Áreas do conhecimento nas quais podem ser aplicados os conceitos da Inteligência Artificial.



**Fonte:** Adaptado de Gomes (2010) e Norvig e Russel (2013)

A Figura 5 demonstra uma representação da IA e suas partições em Aprendizagem de Máquina – *Machine Learning* (ML) e em Aprendizagem Profunda – *Deep Learning* (DL). Por mais que o DL seja uma parte da IA, nesta seção é feita apenas a diferenciação dos grupos para fins de comparação. Deste modo, a aprendizagem de máquina propõe que os sistemas possam aprender a partir dos dados apresentados, extraindo padrões e melhorando seu desempenho de aprendizagem a cada nova hipótese (Ludermir, 2021; Cozman e Kaufman, 2022).

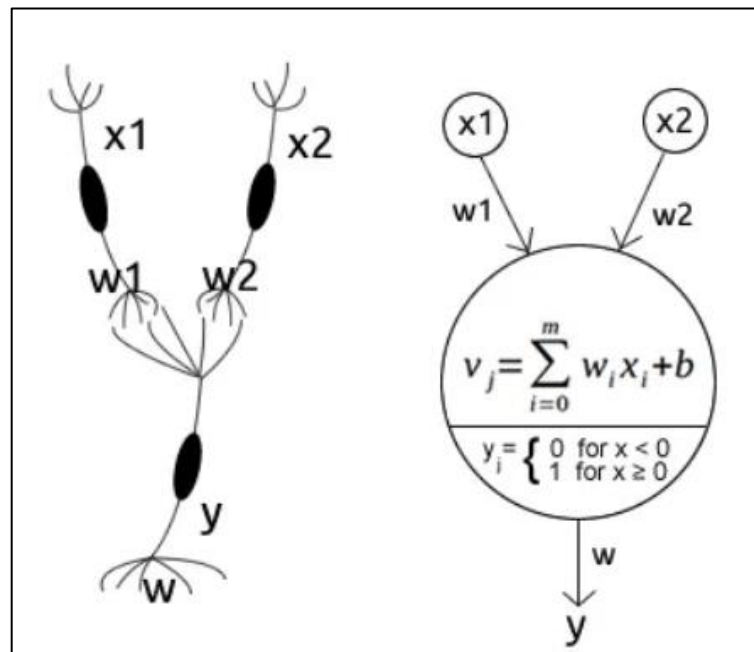
**Figura 5:** Diferenças entre Inteligência Artificial, Aprendizagem de Máquina e Aprendizagem Profunda.



**Fonte:** Adaptado de Khan et al. (2021a)

Além de poderem ser aplicados em problemáticas ambientais, diversos algoritmos dentro do campo da inteligência artificial são bioinspirados, ou seja, se baseiam em processos naturais ou biológicos. Por exemplo, as redes neurais artificiais são inspiradas no funcionamento dos neurônios biológicos (Figura 6), em que um sinal é recebido pelos dendritos ( $x$ ), passado pela célula através do axônio ( $w$ ), e transmitido para outras células por meio da sinapse ( $y$ ).

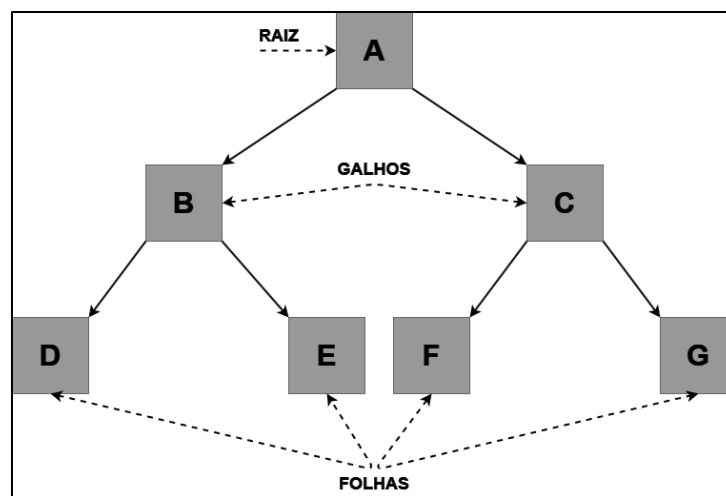
**Figura 6:** Relação entre uma Rede Neural Biológica e uma Rede Neural Artificial



**Fonte:** Grübler (2018)

Outro exemplo é o de algoritmos baseados em árvores, como árvore de decisão, Random Forest e árvore de classificação e regressão. Modelos baseados em árvores são bastante utilizados e abrangem uma grande variedade de problemas, desde os mais simples aos mais complexos. Esses algoritmos se baseiam em criar uma estrutura semelhante a uma árvore (Figura 7), com um nó raiz (A), que representa a primeira pergunta a ser respondida, ramos ou galhos que representam as respostas possíveis de cada teste intermediário (B, C) e, por fim, folhas que representam as decisões finais (D, E, F, G).

**Figura 7:** Funcionamento geral de algoritmos baseados em árvores



**Fonte:** Elaborado pelo Autor (2023)

Ainda existem outros tipos de algoritmos bioinspirados, como algoritmos genéticos, baseados na teoria da evolução, algoritmos baseados em colônias de formigas e de abelhas, algoritmos baseados no sistema imunológico humano e outros baseados no comportamento de pássaros. Cada algoritmo possui seu próprio campo de aplicação.

### 3.6.1 Árvores 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.

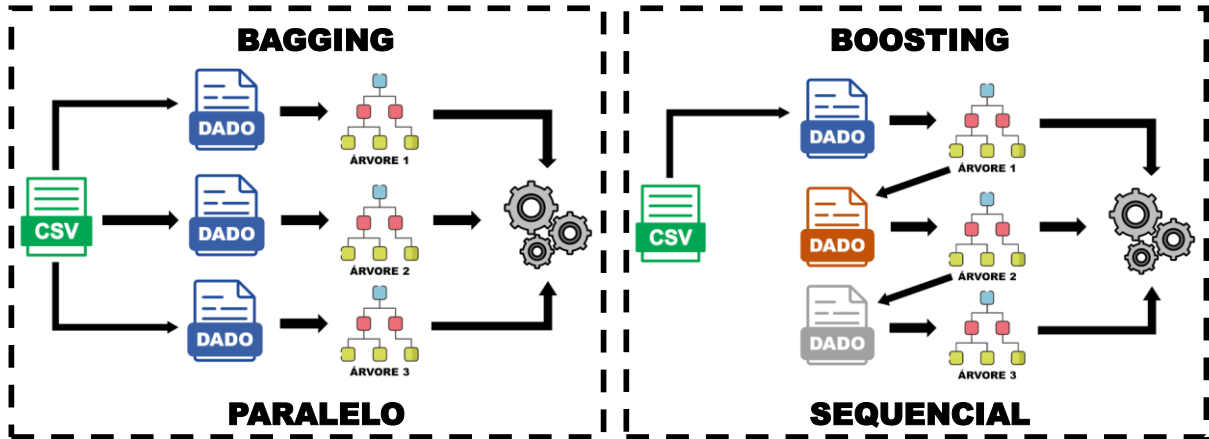
### 3.6.2 Random Forest (RF)

O algoritmo RF é uma evolução das Árvores de Decisão e busca superar algumas de suas limitações, foi introduzido por Ho (1995) e posteriormente desenvolvido 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 (González *et al.*, 2020; Da Silva *et al.*, 2022).

O RF é um método baseado em *bagging* (Breiman, 1996), onde o modelo final faz suas previsões baseado nas previsões de vários modelos treinados de forma

independente (Figura 8), combinadas por meio de votação (em casos de classificação) ou média (para casos de regressão) (Rokach, 2019).

**Figura 8:** Funcionamento de Algoritmos *Ensemble* baseados em *Bagging* e em *Boosting*

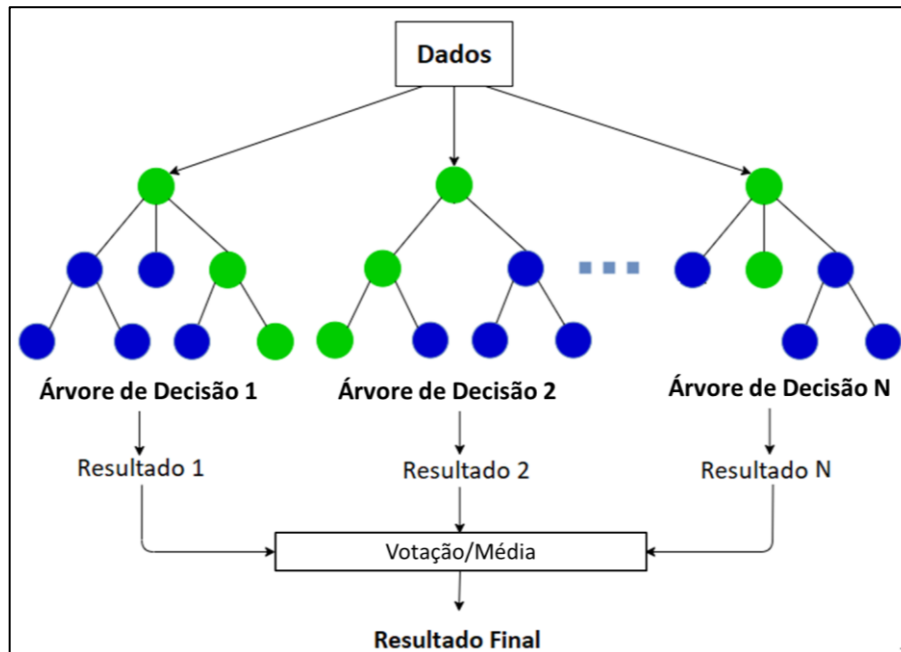


Fonte: Adaptado de Awan (2023)

O funcionamento do RF 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 reduzir a tendência de uma árvore superajustar-se aos dados de treinamento (Mohajane *et al.*, 2021), 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.

Quando se trata de fazer previsões, cada árvore individual do RF emite uma previsão (Figura 9) 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, Theodorisis e Tsadiras, 2022). Além disso, ao contrário das Árvores de Decisão, o RF não requer ajustes complexos de hiperparâmetros, o que facilita seu uso prático.

**Figura 9:** Funcionamento do Processo de Decisão entre as diferentes árvores do Random Forest.



Fonte: Adaptado de Khan *et al.* (2021b)

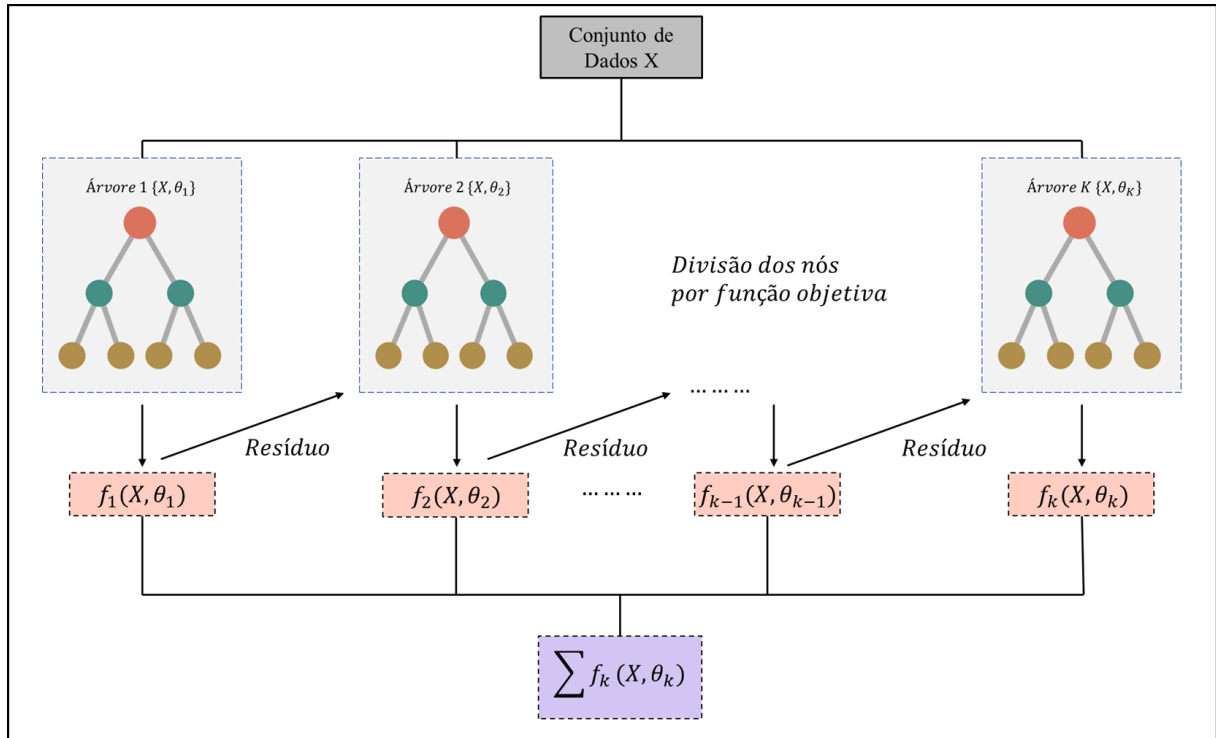
### 3.6.3 Extreme Gradient Boosting (XGB)

Assim como o RF, o XGB é um método *ensemble*, sendo também a combinação de múltiplas árvores de decisão. Diferente do RF, o XGB é baseado no conceito de *boosting* (Schapire e Freund, 2013), em que os modelos são construídos sequencialmente buscando corrigir os erros do modelo anterior (Figura 10). Chen e Guestrin (2016) criaram o XGB como uma implementação específica do método de *Gradient Boosting Tree* (GBT), no qual os erros são minimizados por um algoritmo de gradiente descendente. Como inovações, os autores implementaram diversas otimizações tanto de software (paralelização nas etapas de treinamento, por exemplo), como de hardware (uso mais eficiente do cache de memória, por exemplo), produzindo um modelo mais robusto, eficiente e eficaz do que o modelo mais genérico.

Durante o treinamento, o XGB busca otimizar a função de perda, minimizando a soma do erro residual da previsão anterior, adicionando sequencialmente novas árvores ao modelo. Cada nova árvore busca captar os padrões que não foram aprendidos nos modelos anteriores, utilizando uma função  $f_k(X, \theta_k)$  para ajustar os resíduos da previsão anterior. Ao fim do treinamento, cada amostra de previsão terá uma pontuação correspondente que reflete a contribuição de cada árvore para a

predição final e, então, as pontuações correspondentes são somadas para obter o valor final da predição (Guo *et al.*, 2020).

**Figura 10:** Funcionamento do Processo de Decisão entre as diferentes árvores do XGBoost.



Fonte: Adaptado de Guo *et al.* (2020)

### 3.7 Machine Learning na Predição de Incêndios

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 a quantidades cada vez maiores de dados, chegando aos modelos utilizados atualmente, como CART; RN; SVM; RF, etc.

Da mesma forma que diversos trabalhos têm sido desenvolvidos na área de mapeamento de incêndios florestais, várias pesquisas têm sido feitas aliando técnicas de mapeamento, utilizando Sensoriamento Remoto – SR e Sistemas de Informações Geográficas – SIG, e metodologias de aprendizagem de máquina. O Quadro 2

apresenta alguns trabalhos desenvolvidos nos últimos anos e que utilizam de algoritmos de ML e DL para a predição de incêndios florestais.

Ao longo dos últimos anos, tem crescido consideravelmente a quantidade de pesquisas sendo realizadas utilizando técnicas de ML e de DL para a predição de incêndios florestais ao redor do mundo. Entre os trabalhos realizados destaca-se a grande quantidade de produções acadêmicas na Ásia, como por exemplo o trabalho de Pourghasemi *et al.* (2020a) realizada no Irã. Nesta pesquisa os autores avaliaram a probabilidade de ocorrência de inundações, incêndios florestais e de deslizamentos de terra na Província de Fars, Sudeste do Irã. Para predizer a ocorrência de incêndios, os autores levaram em consideração os 10 fatores mais importantes e submeteram os dados ao modelo RF, obtendo uma acurácia de 95,8%.

Contudo, a grande maioria dos trabalhos de mapeamento da susceptibilidade de incêndios florestais tem sido desenvolvida na China. Pode-se destacar o trabalho de Li *et al.* (2020), que utilizaram os modelos de Máquina de Vetores de Suporte (SVM) e de Redes Neurais para construir modelos de predição de incêndios. Para realizar a predição, os autores utilizaram dados relativos ao fogo, dados meteorológicos, dados do terreno, da vegetação, de infraestruturas e dados socioeconômicos, totalizando 8 dados de entrada e uma resposta caracterizada pela ocorrência ou não de incêndio. Os modelos apresentaram acurácias de 92,16% para a Rede Neural e de 89,89% para a SVM.

A pesquisa de Pang *et al.* (2022) avaliou, por sua vez, quatro algoritmos para predição da ocorrência de incêndios florestais, Rede Neural, *Radial Basis Function Network*, SVM e RF. Os autores avaliaram 32746 focos de incêndio na China entre os anos de 2003 e 2016 e relacionaram esses pontos com dados relativos à ignição do fogo, dados meteorológicos, dados do terreno, de vegetação, de infraestrutura e com dados socioeconômicos. Entre os modelos avaliados, o modelo RF apresentou a maior acurácia, 89,2%, sendo escolhido pelos autores para realizar o mapeamento da probabilidade da ocorrência de incêndios nas áreas estudadas.

**Quadro 3:** Exemplos de Trabalhos de Predição de Incêndios que utilizam de metodologias de IA, ML e DL.

REGIÃO	REFERÊNCIA	ÁREA DE ESTUDO	MODELOS UTILIZADOS
ÁSIA	Pourghasemi <i>et al.</i> (2020a)	Irã	<ul style="list-style-type: none"> <li>• Random Forest</li> </ul>
	Li <i>et al.</i> (2020)	China	<ul style="list-style-type: none"> <li>• Rede Neural</li> <li>• Support Vector Machine</li> </ul>
	Pang <i>et al.</i> (2022)	China	<ul style="list-style-type: none"> <li>• Rede Neural</li> <li>• Radial Basis Function Network</li> <li>• Support Vector Machine</li> <li>• Random Forest</li> </ul>
EUROPA	Michael <i>et al.</i> (2021)	Grécia	<ul style="list-style-type: none"> <li>• Regressão Logística</li> <li>• Random Forest</li> <li>• XGBoost</li> </ul>
	Tonini <i>et al.</i> (2020)	Itália	<ul style="list-style-type: none"> <li>• Random Forest</li> </ul>
	Xie, Peng (2019)	Portugal	<ul style="list-style-type: none"> <li>• Random Forest</li> <li>• Extreme Gradient Boosting</li> <li>• Modelo Linear</li> <li>• Deep Learning</li> </ul>
ÁFRICA	Mohajane <i>et al.</i> (2021)	Marrocos	<ul style="list-style-type: none"> <li>• Rede Neural</li> <li>• Regressão Logística</li> <li>• Árvore de Classificação e Regressão</li> <li>• Random Forest</li> <li>• Support Vector Machine</li> </ul>
	Seddouki <i>et al.</i> (2023)	Marrocos	<ul style="list-style-type: none"> <li>• Random Forest</li> <li>• XGBoost</li> <li>• Support Vector Machine</li> </ul>
AMÉRICAS	Agrawal <i>et al.</i> (2023)	Estados Unidos	<ul style="list-style-type: none"> <li>• Regressão Logística</li> <li>• Árvore de Decisão</li> <li>• Random Forest</li> <li>• Extreme Gradient Boosting</li> <li>• K-Nearest Neighbors</li> <li>• Support Vector Machine</li> </ul>
	Celis <i>et al.</i> (2023)	Colômbia	<ul style="list-style-type: none"> <li>• Rede Neural</li> </ul>
BRASIL	De Bem <i>et al.</i> (2018)	Distrito Federal	<ul style="list-style-type: none"> <li>• Regressão Logística</li> <li>• Rede Neural</li> </ul>
	Juvanhol <i>et al.</i> (2023)	Espírito Santo	<ul style="list-style-type: none"> <li>• Árvore de Classificação e Regressão</li> </ul>
	Rubí <i>et al.</i> (2023)	Distrito Federal	<ul style="list-style-type: none"> <li>• Rede Neural Artificial Profunda</li> <li>• Support Vector Machine</li> <li>• Random Forest</li> <li>• AdaBoost</li> </ul>
	Viganó <i>et al.</i> (2017)	Pantanal	<ul style="list-style-type: none"> <li>• Rede Neural</li> </ul>

Fonte: Elaborado pelo Autor (2024)

Pesquisas também têm sido realizadas na Europa, como é o caso de Michael *et al.* (2021), em que foram avaliados três algoritmos de ML (Regressão Logística, RF e XGB) para o mapeamento do risco de fogo na Grécia. Ao comparar os modelos, observaram que o XGB se mostrou superior aos outros. Tonini *et al.* (2020) também estudaram a utilização do RF para a elaboração de um mapa de susceptibilidade de incêndios florestais na região da Liguria, Itália. Os autores consideraram áreas queimadas ao longo do período entre 1997 e 2017 e 11 fatores geoambientais de predisposição na composição do *dataset* a ser avaliado, tendo apresentado valores aceitáveis, mostrando-se efetivo para lidar com grandes volumes de dados de dimensões espaciais notáveis.

Xie e Peng (2019) utilizaram o conjunto de dados de incêndios florestais da Universidade da Califórnia (UC Irvine), coletados na região nordeste de Portugal, para testar a capacidade de modelos *ensemble* na predição de áreas queimadas por incêndios florestais utilizando modelos de regressão, e também na previsão da ocorrência de incêndios florestais de larga escala utilizando modelos de classificação. No primeiro experimento, ao utilizar modelos de regressão, o algoritmo RF teve um desempenho melhor, tendo apresentado menores valores de RMSE e MAE (*Mean Absolute Error*) em relação aos outros modelos. Por outro lado, entre os modelos de classificação, o método *Extreme Gradient Boosting* teve uma melhor acurácia na predição de incêndios de larga escala.

Alguns trabalhos já têm sido realizados também na África, como é o caso de Mohajane *et al.* (2021) que compararam cinco algoritmos de ML para o mapeamento da suscetibilidade de incêndio florestal no norte do Marrocos, utilizando 510 pontos históricos de incêndios como inventário, além de 10 fatores independentes, entre eles uso da terra, elevação do terreno, temperatura e o índice de vegetação por diferença normalizada (NDVI). Os modelos apresentados mostraram uma acurácia entre 80 e 98,9%, mostrando que a utilização de modelos híbridos pode melhorar consideravelmente a performance de estudos sobre a suscetibilidade de incêndios florestais.

Seddouki *et al.* (2023) também avaliaram a susceptibilidade de incêndios florestais no norte do Marrocos utilizando os algoritmos RF, XGB e SVM. Os autores utilizaram como variável alvo as ocorrências de fogo no período entre janeiro de 2008 e maio de 2021 e, como fatores de explicação, um conjunto de 10 variáveis: elevação,

declividade, orientação de vertentes, uso e cobertura da terra, NDVI, precipitação, velocidade do vento e a distância de estradas e residências. Os resultados de validação apontaram que o modelo XGB teve um melhor desempenho em relação aos outros dois modelos. Por outro lado, pelos valores apresentados, todos os algoritmos tiveram desempenho satisfatório na identificação de áreas susceptíveis a ocorrência de incêndios florestais.

Na pesquisa realizada por Agrawal et al. (2023), os autores utilizaram seis modelos de classificação para prever a ocorrência de incêndios de larga escala. Foram utilizados dados referentes a 2109 incêndios ocorridos nos Estados Unidos no período entre 1992 e 2020, enquanto o conjunto de variáveis preditoras foi constituído por 38 variáveis, sendo cinco variáveis atmosféricas avaliadas em quatro diferentes pressões e as médias dos três anos anteriores a cada ocorrência de seis produtos do sensor MODIS: o NDVI, o Índice Melhorado de Vegetação (EVI - *Enhanced Vegetation Index*), o índice de área foliar (LAI - *Leaf Area Index*), Fração da Radiação Fotossinteticamente Ativa (FPAR - *Fraction of Photosynthetically Active Radiation*) e as temperaturas de superfície durante o dia (LST Day) e durante a noite (LST Night). Dentre os algoritmos testados, os modelos XGB e RF se destacaram dos demais, com acurácias de 90.44% e 87.62%, respectivamente.

Em um contexto mais próximo, Celis et al. (2023) conduziram um estudo na Amazônia colombiana com o objetivo de identificar as condições específicas meteorológicas, de uso e cobertura do solo e de fatores socioeconômicos predominantes sobre a ocorrência de incêndios florestais. Os Autores utilizaram redes neurais com camadas do tipo *Long-Short Term Memory* (LSTM), obtendo resultados consistentes quanto ao desempenho para representar tanto a magnitude quanto a evolução temporal da variável alvo.

Contudo, apesar de já existir uma grande quantidade de estudos realizados nesta temática mundo afora, a quantidade de trabalhos realizados no Brasil ainda é incompatível com a grande importância ambiental dos diversos biomas presentes no território nacional, especialmente a Amazônia Brasileira. Ainda assim, alguns trabalhos têm sido desenvolvidos principalmente nas regiões sudeste e centro-oeste, como a pesquisa de De Bem et al. (2018), que compararam os modelos de Regressão Logística e de Rede Neural Artificial para predição da ocorrência de fogo no Distrito Federal. Os resultados mostraram que os modelos tiveram desempenhos similares,

com uma pequena vantagem para a Rede Neural, mostrando que esses métodos podem oferecer bons resultados a partir dos dados disponíveis no país.

Outro trabalho realizado no Distrito Federal foi o de Rubí *et al.* (2023) que consideraram 19 variáveis relacionadas à ocorrência de incêndios através de quatro modelos: RNP, SVM, RF e *AdaBoost*. Ao compararem as métricas de cada modelo, os autores apresentam os modelos *AdaBoost* e RF com desempenhos melhores em relação aos demais, confirmando a ideia de que ferramentas de ML podem prever com qualidade uma área queimada a partir de dados topográficos, climáticos, hidrográficos e antropogênicos.

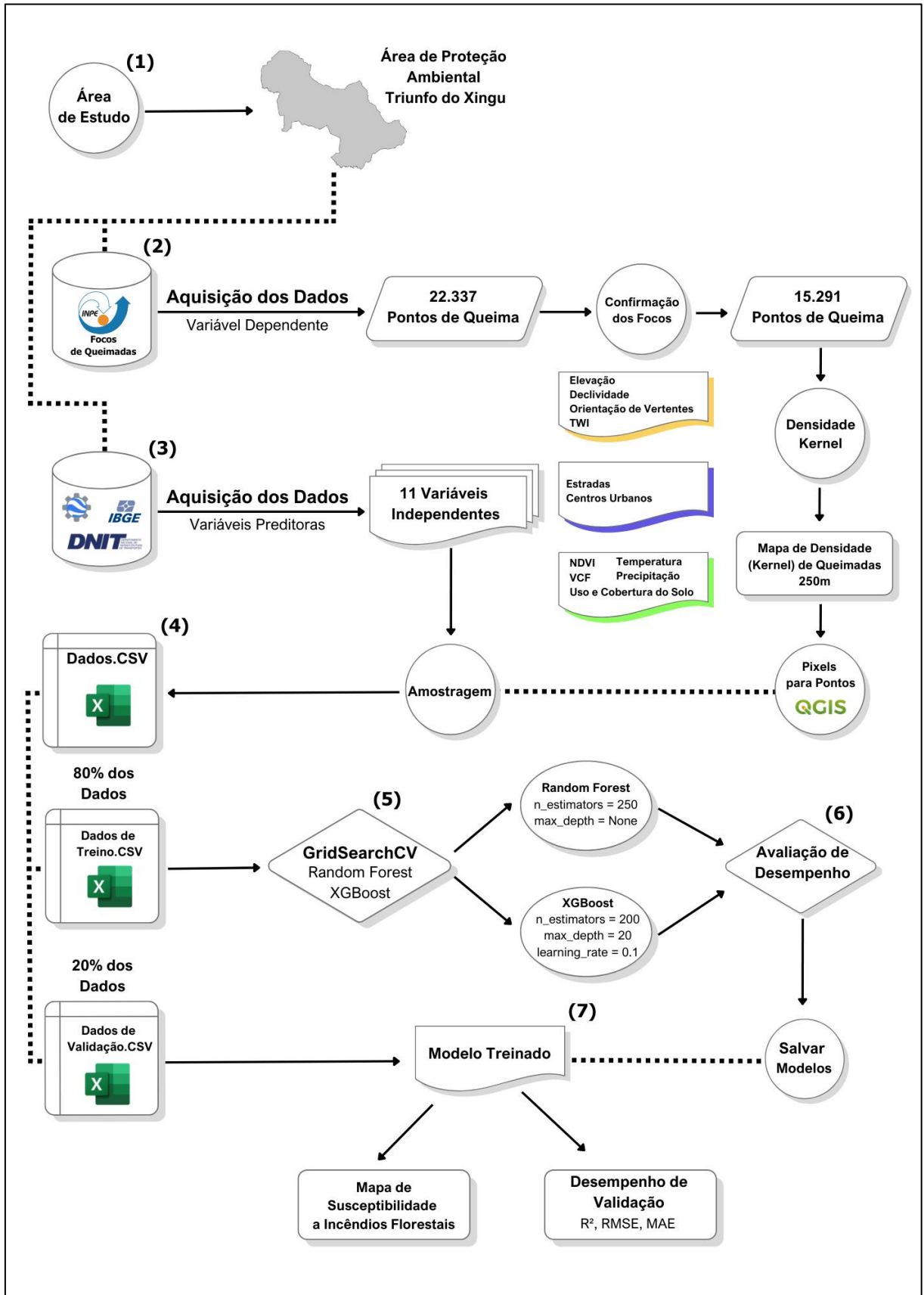
O pantanal também é uma área com grande importância ambiental ainda muito carente de estudos sobre a susceptibilidade de incêndios. Uma das pesquisas realizadas nessa área é a de Viganó *et al.* (2017), que tinham como objetivo aplicar a técnica de Redes Neurais para a previsão de focos e de área queimada no Mato Grosso do Sul. Como resultados, os autores obtiveram valores de  $R^2$  de 0,903 utilizando um conjunto de variável meteorológicas como preditoras e de 0,986 utilizando somente os dados de área queimada.

Ainda assim, a grande maioria dos trabalhos científicos realizados no Brasil se concentram na região Sudeste do país. Como exemplo de trabalhos realizados nessa região temos Juvanhol *et al.* (2023) que para obter um mapa de ocorrência de fogo para o estado do Espírito Santo, utilizaram o modelo de árvore de classificação e regressão (CART), a partir de 12 variáveis preditoras relativas a um período de 16 anos (2000 a 2015). Os resultados mostraram que a combinação dos modelos de ML e as técnicas de Sistemas de Informações Geográficas puderam produzir modelos de predição significantes, além de confirmar a grande importância dos fatores socioeconômicos, ambientais e de vegetação para os modelos propostos.

#### **4 MATERIAL E MÉTODOS**

O fluxograma relativo à metodologia é apresentado na Figura 11. Serão apresentados os dados utilizados, sua aquisição, pré-processamento, bem como a aplicação e comparação entre os modelos estudados e por fim, a elaboração do mapa de susceptibilidade de incêndios.

**Figura 11:** Representação Visual dos Principais Procedimentos e Ferramentas Envolvidos



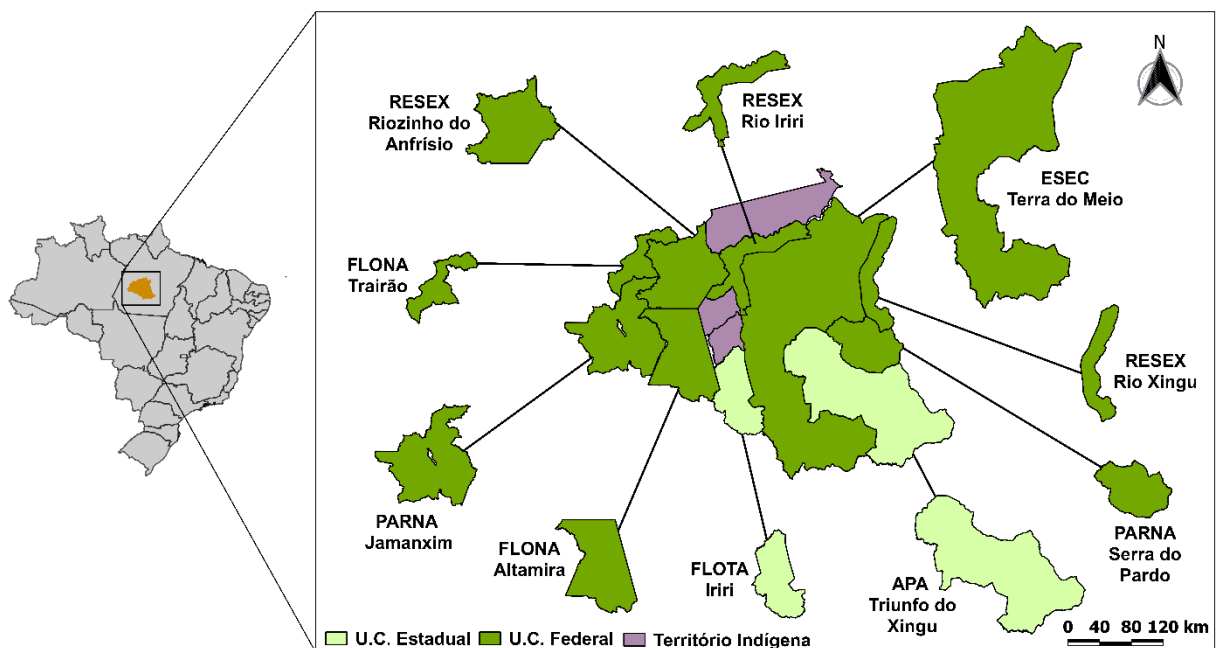
Fonte: Elaborado pelo Autor (2023)

#### 4.1 Área de Estudo

A área de estudo é representada pela Área de Preservação Ambiental (APA) Triunfo do Xingu, uma unidade de conservação de uso sustentável localizada no estado do Pará, criada pelo Decreto Estadual nº 2.612, em 04 de dezembro de 2006. A APA possui uma área total de 1.679.280,52 hectares, distribuídos entre os municípios de Altamira e São Félix do Xingu. O Sistema Nacional de Unidades de Conservação (SNUC), instituído pela Lei Federal nº 9.985, de 18 de julho de 2000, estabelece que unidades de uso sustentável tenham como objetivo principal compatibilizar a conservação da natureza e o uso sustentável de seus recursos.

A APA Triunfo do Xingu faz parte de um mosaico de unidades protegidas denominado Terra do Meio (Figura 12), composto por ela e outras 9 unidades de conservação (UC's): Reservas Extrativistas (RESEX) Riozinho do Anfrísio, Rio Iriri e Rio Xingu; Parques Nacionais (PARNA) Serra do Pardo e do Jamanxim; Estação Ecológica (ESEC) Terra do Meio; Florestas Nacionais (FLONA) Trairão e Altamira e a Floresta Estadual (FLOTA) do Iriri. Além das 10 unidades de conservação, o Mosaico da Terra do Meio também é composto por Territórios Indígenas (TI) localizados na região.

**Figura 12:** Unidades de Conservação Federais, Estaduais e Territórios Indígenas que compõem o Mosaico de Áreas Protegidas Terra do Meio

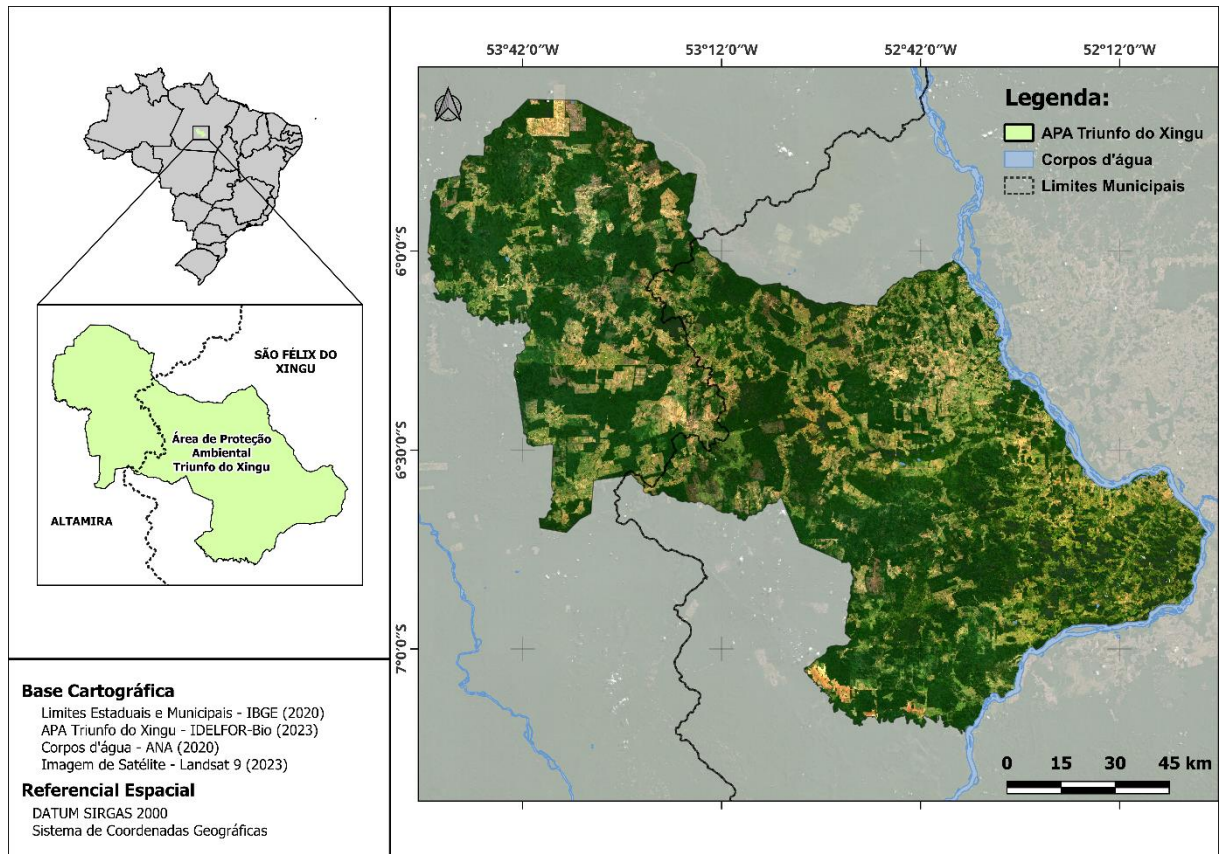


Fonte: Elaborado pelo Autor (2024)

Ainda assim, a APA Triunfo do Xingu já tem sido apontada por diversos estudos como uma área crítica em relação ao desmatamento e à ocorrência de incêndios, principalmente, como apontado por Araújo *et al.* (2017).

A APA Triunfo do Xingu está localizada na mesorregião do Sudeste Paraense (Figura 13), onde o clima é caracterizado como tropical úmido, categoria Ami, segundo a classificação climática de Köppen, com duas estações bem definidas, uma relativamente seca, entre os meses de maio e outubro e outra chuvosa, entre os meses de novembro e abril. A região tem uma precipitação média anual de 2000 mm.ano<sup>-1</sup>, uma temperatura média anual de 27 °C e umidade relativa média de 80% (Carvalho *et al.*, 2022).

**Figura 13:** Localização Geográfica da Área de Estudo, a APA Trinfo do Xingu, no Estado do Pará, Brasil

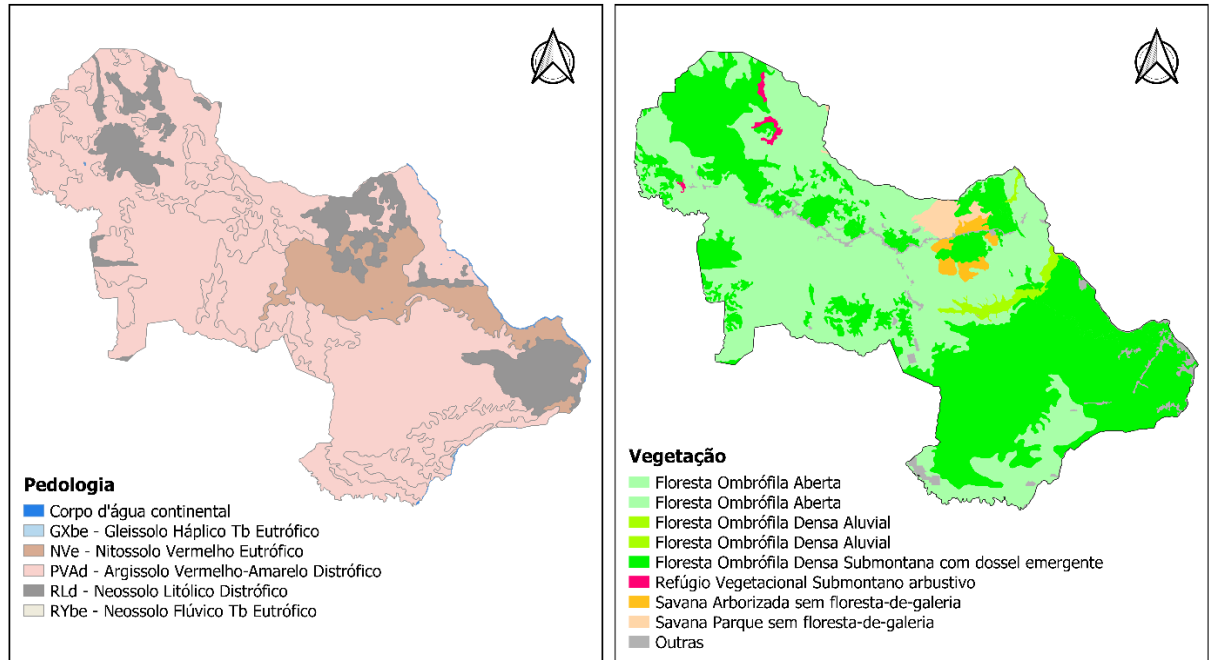


**Fonte:** Elaborado pelo Autor (2023)

A vegetação da região apresenta uma predominância de Florestas Ombrófilas, que ocupam grande parte da região, contudo existem algumas áreas de transição Floresta-Cerrado, onde predomina vegetação de Savana. Os solos mais comuns na região (Santos *et al.*, 2018) são os Argissolos Vermelho-Amarelo, Nitossolos

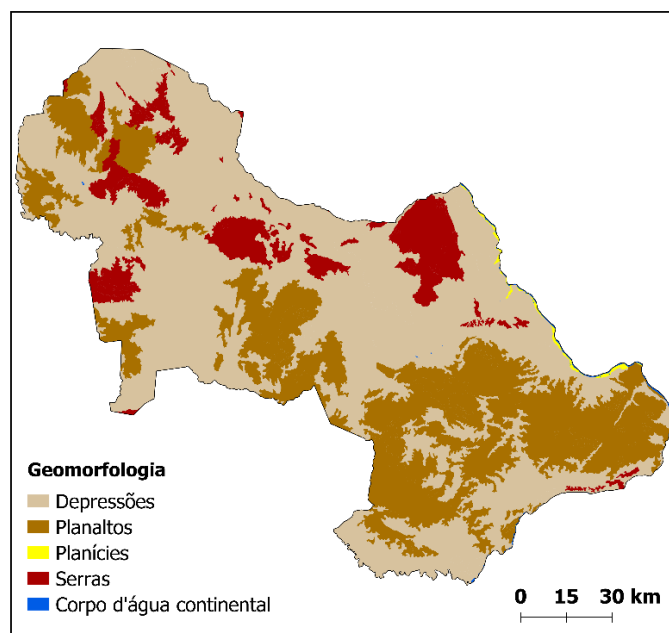
Vermelho e Neossolos Litólicos, havendo também a presença de Gleissolos Háplicos e Neossolos Flúvicos (Figura 14).

**Figura 14:** Caracterização dos Solos e da Vegetação da Área de Proteção Ambiental Triunfo do Xingu, no Estado do Pará, Brasil.



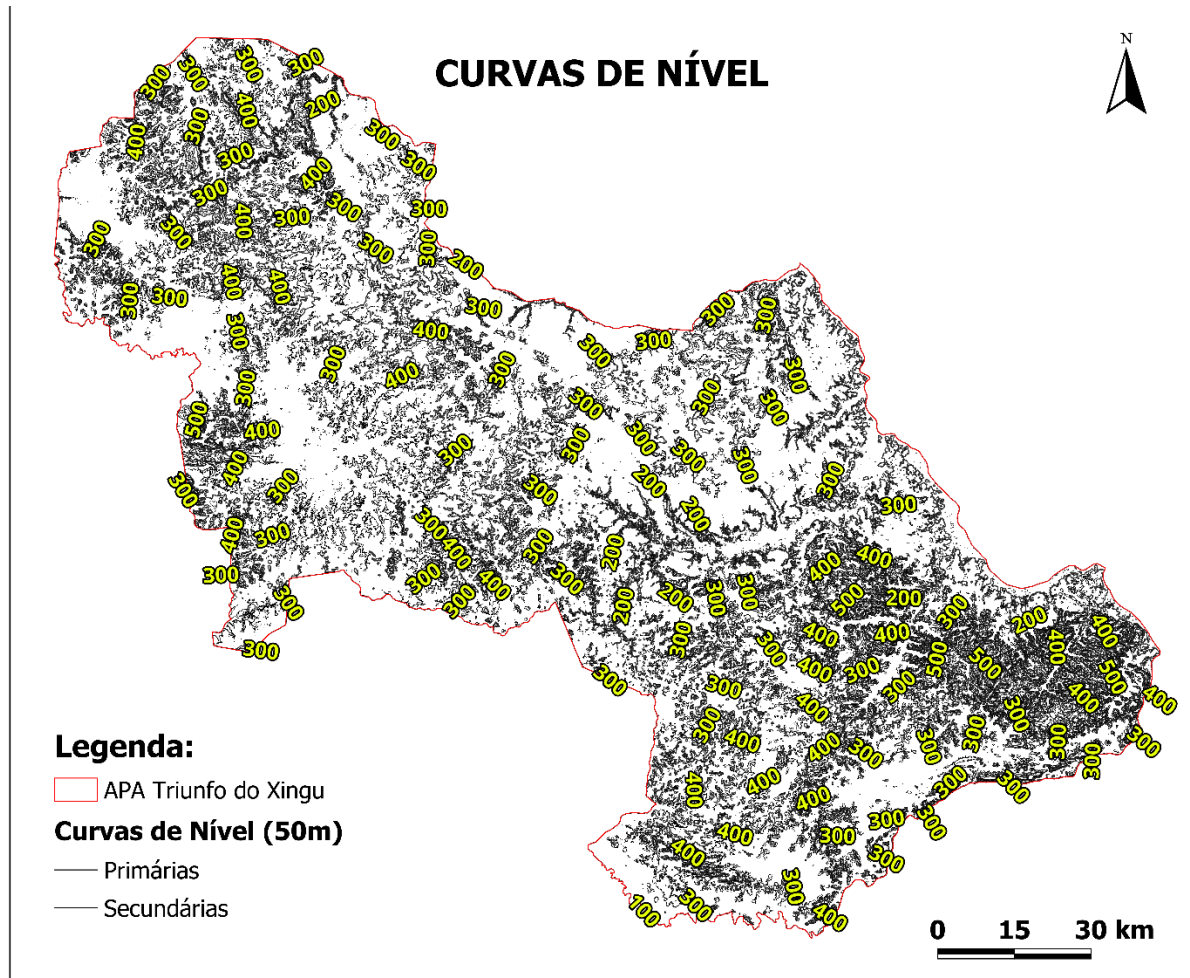
Fonte: Elaborado pelo Autor (2023)

**Figura 15:** Caracterização Geomorfológica da Área de Proteção Ambiental Triunfo do Xingu, no Estado do Pará, Brasil.



Fonte: Elaborado pelo Autor (2023)

**Figura 16:** Mapa de Curvas de Nível da Área de Proteção Ambiental Triunfo do Xingu, no Estado do Pará, Brasil.



Fonte: Elaborado pelo Autor (2023)

#### 4.2 Densidade (Kernel) de Queimadas – Variável Resposta

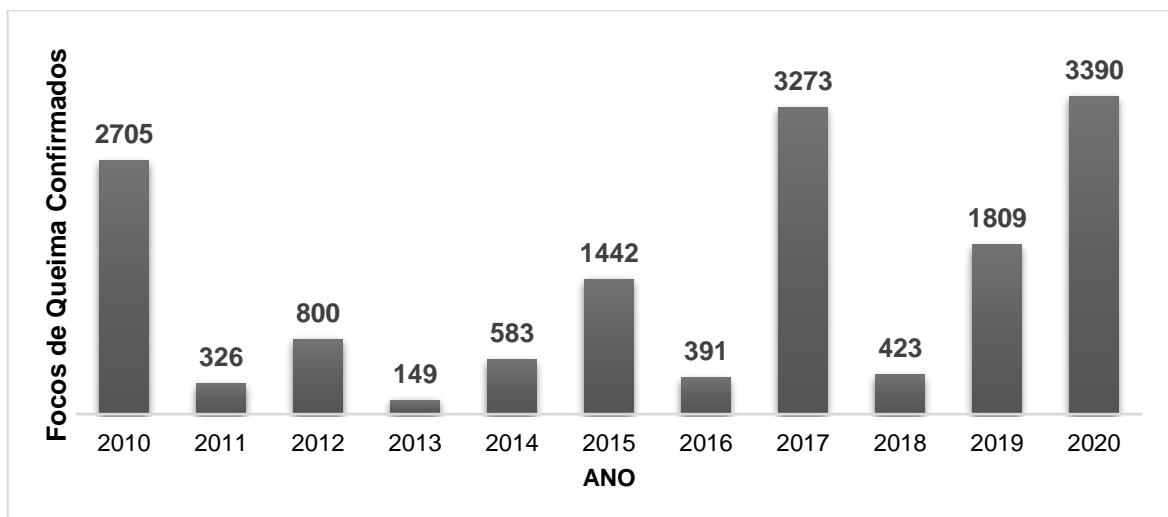
Como variável alvo ou resposta, foi utilizada a ocorrência de focos de queima, representada pela densidade de focos. Os dados referentes aos focos de queimadas são disponibilizados para download pelo Programa Queimadas (BDQueimadas) INPE, hospedado no sítio <https://terrabilis.dpi.inpe.br/queimadas/portal/>. Cada foco indica a ocorrência de fogo em um pixel de tamanho variável, dependendo da resolução espacial do satélite.

O atual satélite de referência utilizado pelo INPE é o AQUA\_M-T, que possui resolução espacial de 1 km e que é capaz de detectar frentes de fogo com, pelo menos, 30 metros de extensão por 1 metro de largura. Dessa forma, em um único pixel podem haver uma ou várias frentes de fogo, gerando apenas um foco com vários

eventos. Por outro lado, um único evento muito extenso pode ser detectado em diferentes pixels e, dessa forma, gerando múltiplos focos para um evento único (INPE, 2023).

Para este trabalho, os dados foram filtrados para a área interna da APA Triunfo do Xingu, sendo obtidos ano a ano para o período de 2010 a 2020, totalizando um total de 22.337 focos, detectados pelo satélite de referência. Contudo, uma vez que os focos não representam necessariamente um evento de queima, é necessário realizar a confirmação desses pontos. Para isso, utilizou-se o produto de Áreas Queimadas, também disponibilizado pelo INPE (Libonati *et al.* 2015; INPE, 2023), validado no trabalho de Rodrigues *et al.* (2019). Foram baixados os dados mensais de áreas queimadas para o período de 2010 a 2020 e comparados com os dados pontuais baixados anteriormente, de forma que, apenas aqueles pontos que estivessem contidos em áreas queimadas foram mantidos na análise. Por fim, restaram 15.291 pontos de queima confirmados na área da APA Triunfo do Xingu, conforme representado na Figura 15.

**Figura 17:** Quantidade de focos de queimada registrados pelo Instituto Nacional de Pesquisas Espaciais (INPE) confirmados na APA Triunfo do Xingu entre os anos de 2010 e 2020



Fonte: INPE/BDQueimadas (2023)

A metodologia utilizada para realizar a estimativa de Kernel seguiu a proposta de Rizzatti *et al.* (2020), na qual são necessários dois fatores principais, o raio de influência (R) e a função de estimação (k). O raio de influência irá determinar a área centrada em torno do ponto que se deseja estimar em que estarão os pontos que

contribuirão para a estimativa, enquanto a função de estimação estará relacionada com as propriedades para suavização do fenômeno.

Para calcular a densidade Kernel, os arquivos vetoriais de cada ano foram importados para o software QGIS 3.28.15, reprojitados para um sistema de coordenadas planas e unidos em uma única camada. Dessa forma, o raio de influência (R) pode ser calculado a partir da seguinte equação:

$$R = \bar{X} \pm \bar{X}_\sigma \quad (1)$$

Onde,  $\bar{X}$  representa a média da distância média entre os focos e  $\bar{X}_\sigma$  representa a média do desvio padrão.

Para calcular a média das distâncias médias e dos desvios padrão, foi utilizada a ferramenta “Matriz de Distância” que retorna a distância média, o desvio padrão, a distância máxima e mínima de cada ponto e então são calculadas as médias desejadas. Dessa forma, o raio que melhor proporcionalizou foi o da subtração das médias e a função de estimação selecionada foi a função quártica, por atribuir peso maior aos pontos mais próximos em detrimento dos mais distantes, mas apresentando um decaimento gradual conforme a distância aumenta (Porter e Reich, 2012). Além disso, foi determinada a resolução da camada raster de saída, a partir do tamanho dos pixels, como 250 metros.

Para determinar o dataset representativo de toda a área utilizou-se a ferramenta Pixels de Raster para Pontos, transformando cada um dos pixels da imagem em pontos que servirão como base para extrair as informações de cada uma das variáveis independentes. Ou seja, com um ponto de coleta a cada 250 metros.

### 4.3 Variáveis Predictoras

As variáveis predictoras utilizadas estão destacadas no quadro 3, sendo elas: Elevação, Declividade, Orientação de Vertentes, Índice Topográfico de Umidade (TWI – *Topographical Wetness Index*), Precipitação, Temperatura, Proximidade de Centros Urbanos, Proximidade de Estradas, Uso e Cobertura da terra, Campo Contínuo de Vegetação, o Índice de Vegetação por Diferença Normalizada (NDVI – *Normalized Difference Vegetation Index*).

**Quadro 4:** Variáveis consideradas como preditoras na análise dos modelos de Predição de Incêndios

	VARIÁVEL	ABREVIÇÃO	FONTE	RESOLUÇÃO	ALCANCE
<b>TOPOGRÁFICAS</b>	Elevação	ELEVATION			167 m → 644 m
	Declividade	SLOPE	Modelo Digital de Elevação (SRTM)	30 metros	0% → 60%
	Orientação de Vertentes	ASPECT			0° → 360°
	Índice Topográfico de Umidade	TWI			7,3 → 17,06
<b>SOCIOECONÔMICAS</b>	Distância de Estradas	DIST_ESTRA	DNIT	-	1 km → 15 km
	Distância de Áreas Habitadas	DIST_URB	IBGE	-	1 km → 100 km
<b>AMBIENTAIS</b>	Campo Contínuo de Vegetação	VCF	MODIS (MOD44B)	250 metros	8% → 81%
	Temperatura	TEMP	LANDSAT 7 8	30 metros	27 °C → 42 °C
	Precipitação	PRECIPITATION	CHIRPS	0.05°	2000 → 2300 mm.ano <sup>-1</sup>
	Índice de Vegetação por Diferença Normalizada	NDVI	LANDSAT 7 8	30 metros	-1 → 1
	Uso e Cobertura do Solo	USO_SOLO	Mapbiomas Coleção 8	30 metros	Floresta Savana Campos Alagados Formação Campeste Pastagem Lavoura Temporária Soja Mineração Rios e Lagos

**Fonte:** Elaborado pelo Autor (2024)

As subseções a seguir descrevem as metodologias utilizadas para a aquisição de cada conjunto de variáveis, divididos em três grupos: variáveis topográficas, variáveis socioeconômicas e variáveis ambientais.

#### 4.3.1 Variáveis Topográficas

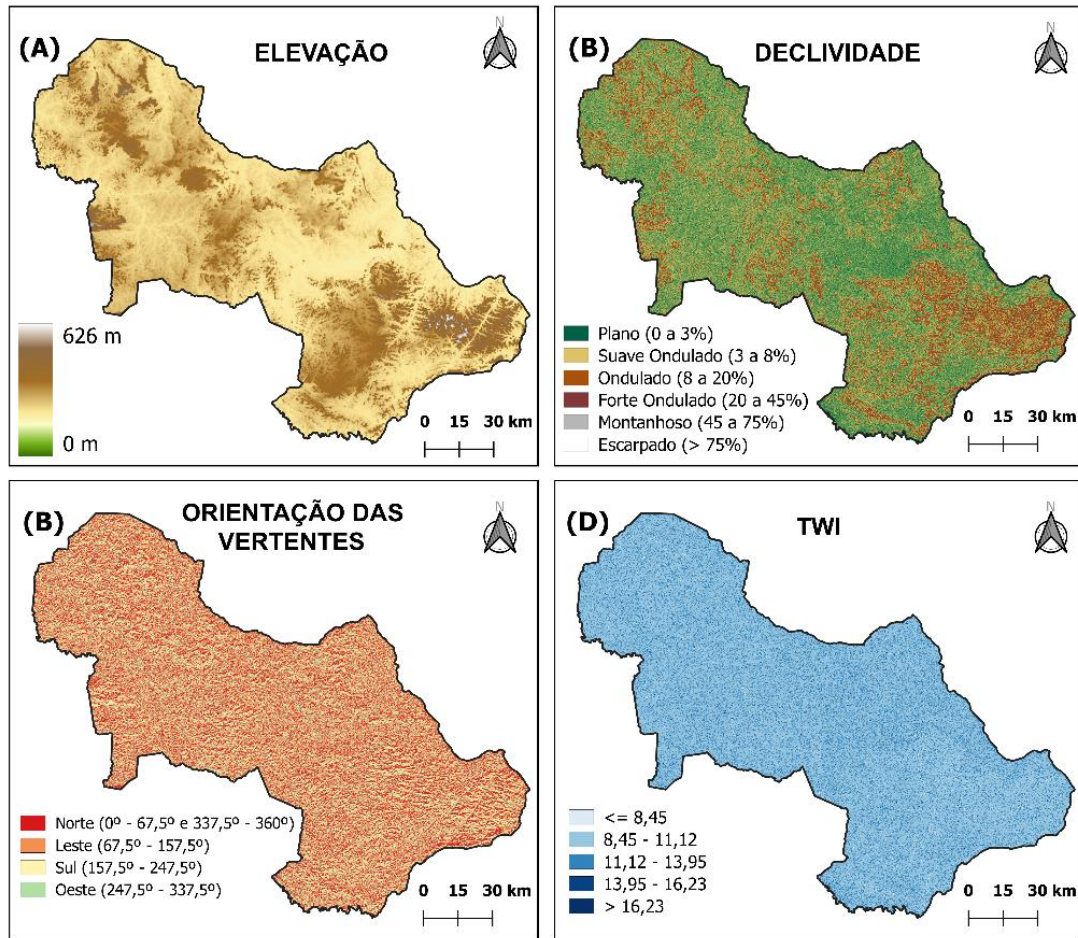
Os subconjuntos de dados topográficos: Elevação, Declividade, Orientação de Vertentes e o TWI (Figura 16), foram obtidos com o auxílio da ferramenta *Google Earth Engine* – GEE, derivados do Modelo Digital de Elevação – MDE elaborado a partir dos dados da SRTM (*Shuttle Radar Topographic Mission*) disponibilizados pelo USGS (*United States Geological Survey*), com resolução de 30 metros.

O Índice Topográfico de Umidade (TWI – *Topographic Wetness Index*) reflete a tendência do acúmulo de água em um determinado ponto do terreno, de modo que quanto menor o valor, menor é o acúmulo de água na área e, portanto, uma maior probabilidade da ocorrência de fogo (Pourtaghi *et al.*, 2015; Nóbrega *et al.*, 2018). Por definição, o TWI é calculado por meio da equação, inicialmente proposta por Beven e Kirkby (1979):

$$TWI = \ln\left(\frac{\alpha}{\tan \beta}\right) \quad (2)$$

Onde,  $\alpha$  representa a área de contribuição e  $\beta$  é o valor da declividade em radianos.

**Figura 18:** Características de (A) Elevação, (B) Declividade, (C) Orientação de Vertentes e (D) Índice Topográfico Composto, para a área da APA Triunfo do Xingu



**Fonte:** Elaborado pelo Autor (2024)

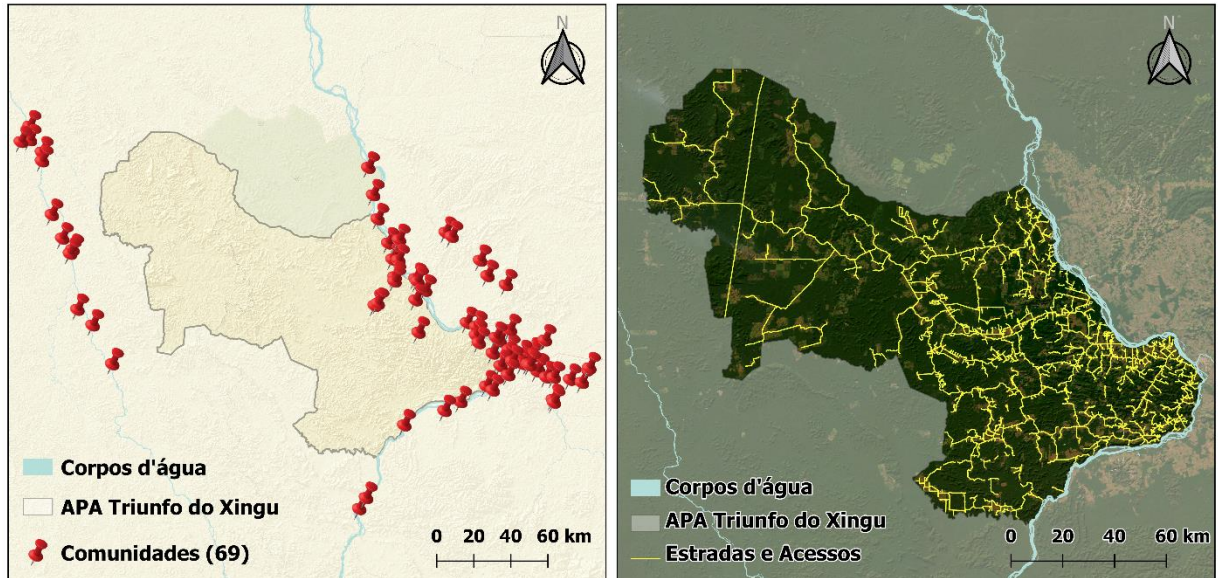
#### 4.3.2 Variáveis Socioeconômicas

As variáveis associadas à presença humana e às atividades socioeconômicas, representadas pela proximidade a centros e localidades urbanas e de estradas (Figura 17) foram obtidas a partir da ferramenta *Multi Ring Buffer* disponível no *software* QGIS 3.28.7, que cria uma série de buffers em distâncias especificadas em torno das localidades e das estradas e, então esses buffers são mesclados e dissolvidos para evitar a sobreposição dos valores.

Os dados originais de localidades e centros urbanos foram provenientes do Instituto Brasileiro de Geografia e Estatística – IBGE (2010). Por conta da metodologia empregada quanto a distância, foram consideradas comunidades distantes até cinco quilômetros da APA. As estradas foram obtidas por meio da correção manual dos

arquivos vetoriais *shapefile* de rodovias federais e estaduais disponibilizados pelo Departamento Nacional de Infraestrutura de Transportes – DNIT (2022).

**Figura 19:** Comunidades e Estradas Localizadas na APA Triunfo do Xingu e no seu entorno



**Fonte:** Elaborado pelo Autor (2024)

#### 4.3.3 Variáveis Ambientais

Para representar o subconjunto de variáveis associadas ao meio ambiente em si foram utilizadas as seguintes características: classes de Uso e Cobertura da terra, Campo Contínuo de Vegetação e o Índice Normalizado de Vegetação (*Normalized Difference Vegetation Index* – NDVI) (Figura 18).

As classes de uso e cobertura da terra são produtos da coleção 8 da Série Anual de Mapas de Cobertura e Uso da Terra, que inclui o período de 1965 a 2022, do projeto MapBiomass (MapBiomass Project, 2023; Souza *et al.*, 2020). As classes foram baixadas com o auxílio do GEE em formato GeoTIFF na escala de 1:100.000 e uma resolução espacial de 30 metros, utilizando o ano de 2020 como referência.

O índice de vegetação por diferença normalizada (NDVI) foi calculado a partir das imagens Landsat com resolução espacial de 30 metros, envolvendo o período de 2010 a 2020. No pré-processamento das imagens foram realizadas correções radiométricas referentes às diferentes gerações dos sensores Landsat envolvidos (Landsat 7 e 8), removidos os agentes interferentes (nuvens e sombras, consideradas até 10%), calculada a média anual dos valores de NDVI para cada um dos anos de

interesse e, finalmente, extraída a mediana desses valores e gerada uma única imagem representativa para todo o período. O NDVI pode ser calculado pixel a pixel a partir da seguinte equação:

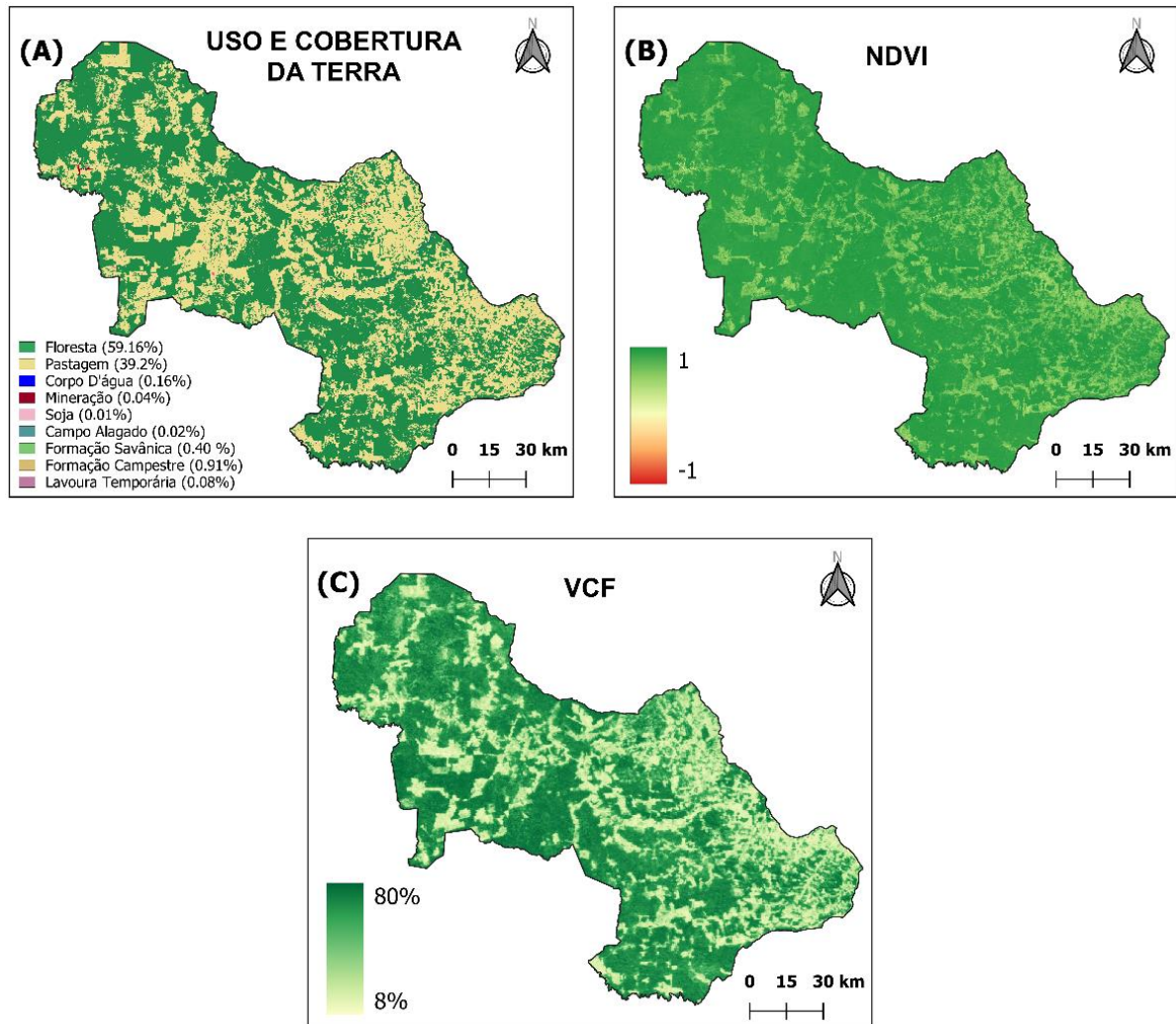
$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (3)$$

Onde NIR representa os valores de reflectância na banda do infravermelho próximo e RED representa os valores de reflectância na banda no vermelho.

A variável de Campo Contínuo de Vegetação (VCF) é representada pelo componente de porcentagem de cobertura de vegetação (*percent tree cover*) do produto anual MODIS *Vegetation Continuous Fields* (MOD44B), com uma resolução espacial de 250 metros. As imagens foram obtidas com o auxílio do Google Earth Engine para cada ano do período de 2005 a 2020 e, por fim, foi extraída a mediana em cada um dos pixels para representar a totalidade do período.

Os fatores climáticos, temperatura e precipitação, influenciam diretamente a ocorrência, a frequência e a intensidade de incêndios, facilitando ou retardando a ignição do combustível (Vadrevu *et al.*, 2010; Moreira *et al.*, 2020). Desse modo, os dados relativos à precipitação foram obtidos pelo GEE, a partir do conjunto de dados *Climate Hazards Group Infrared Precipitation with Station data* (CHIRPS), que possui dados diários a cada três horas e com uma resolução espacial de aproximadamente 0.05°. Neste estudo, foi calculada a média da precipitação anual para os anos de 2010 a 2020 (Figura 19).

**Figura 20:** Características de (A) Uso e Cobertura da Terra, (B) Índice de Vegetação por Diferença Normalizada e (C) Campo Contínuo de Vegetação, da APA Triunfo do Xingu



Fonte: Elaborado pelo Autor (2024)

Para estimar a Temperatura de Superfície (LST) utilizou-se a metodologia proposta por De Jesus e Santana (2017) para o cálculo da temperatura a partir de imagens Landsat. A temperatura média (Figura 19) levou em consideração o período de 2010 a 2020, de forma que, a Temperatura de Superfície em °C foi calculada pixel a pixel para cada um dos anos pela seguinte equação:

$$LST = \frac{T_{\beta}}{1 + \left( \lambda \times \frac{T_{\beta}}{1,438} \times \ln \varepsilon \right)} - 273,15 \quad (4)$$

Onde,  $T_{\beta}$  representa o valor da Temperatura de Brilho ou Temperatura *at-satellite*, calculada pela conversão da banda termal em número digital pelos fatores de escala

e offset,  $\lambda$  representa o comprimento de onda da banda termal, e  $\varepsilon$  representa a emissividade calculada pela seguinte equação:

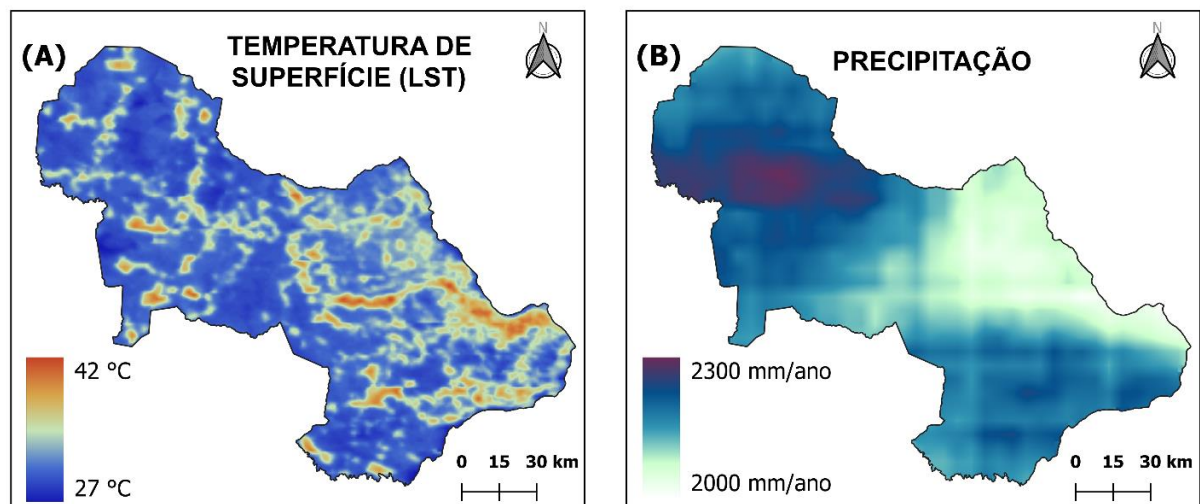
$$\varepsilon = 0,004 \times fv + 0,986 \quad (5)$$

Onde  $fv$  representa a fração de vegetação, que é obtida pela equação:

$$fv = \left( \frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}} \right)^2 \quad (6)$$

Onde, NDVI representa o índice de vegetação por diferença normalizada, obtido pela equação 3,  $NDVI_{min}$  representa o menor NDVI apresentado e  $NDVI_{max}$ , o maior valor de NDVI.

**Figura 21:** Características de (A) Temperatura de Superfície e (B) Precipitação Média Anual, da APA Triunfo do Xingu



Fonte: Elaborado pelo Autor (2024)

#### 4.4 Seleção de Variáveis

No conjunto inicial de variáveis preditoras, além das 11 variáveis que foram utilizadas neste estudo, foram consideradas também outras duas: o Índice Melhorado de Vegetação (EVI – *Enhanced Vegetation Index*) e o Índice de Umidade por Diferença Normalizada (NDMI – *Normalized Difference Moisture Index*).

Contudo, um dos maiores desafios para a seleção de variáveis é evitar o problema da multicolinearidade, ou seja, a presença de variáveis que tenham uma alta correlação entre si, o que pode tornar o modelo final questionável em relação a sua capacidade de predição (Harrison, 2020; Handorf *et al.*, 2020). Entre os efeitos da

multicolinearidade estão a queda na precisão e na capacidade de generalização do modelo (Chan *et al.*, 2022). Neste caso, para determinar a correlação entre as variáveis preditoras foram calculados os coeficientes de correlação de Spearman.

## 4.5 Processamento dos Dados

### 4.5.1 Pré-Processamento

Os dados foram pré-processados em ambiente de programação utilizando a linguagem *Python* (<https://www.python.org/>) com o auxílio da biblioteca *pandas* (<https://pandas.pydata.org/>), que possibilita a análise e a manipulação de dados. Neste caso foi verificado se haviam dados ausentes que precisariam ser substituídos, no caso de valores qualitativos, pelos valores de maior frequência e, no caso de valores quantitativos, pela média (Batista e Monard, 2003).

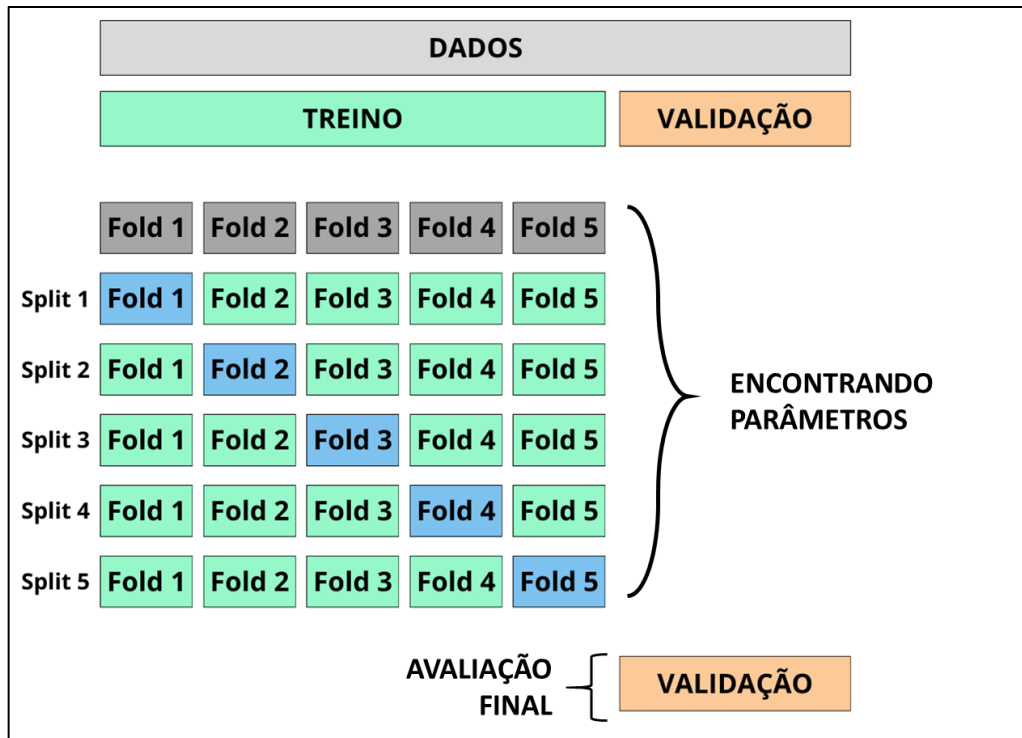
Neste caso particular de problema de regressão, os valores de entrada precisam ser numéricos. Dessa forma, valores de entrada categóricos são substituídos por valores quantitativos. As duas principais formas de realizar essa troca são o *Integer Encoding*, no qual se atribui um valor inteiro único para cada variável categórica, de forma que cada categoria terá um número que a represente, podendo iniciar a listagem em Zero ou Um. E a outra forma seria o *One-hot Encoding*, em que cada categoria dará origem a uma nova coluna binária – por exemplo: se existem cinco classes de uso e cobertura do solo, serão criadas cinco novas colunas, uma para cada classe, que substituirão a anterior. Neste caso, optou-se por utilizar o *Integer Encoding*, evitando a criação de novas colunas de informação, de modo a preservar o número de entradas nos modelos utilizados.

### 4.5.2 Validação Cruzada – Dados de Treino e de Validação

Os dados foram divididos em conjuntos de treino/teste e validação, com a proporção de 80% dos dados para treinar e testar os modelos e os outros 20% para validar os modelos já treinados. A divisão dos dados em conjuntos de treino/teste e de validação tem como objetivo testar a capacidade de generalização dos modelos e avaliar o quão bem o aprendizado do modelo se aplica a dados que não foram usados nas etapas de treinamento (Dangeti, 2017).

Para efeitos de comparação entre os modelos testados, foi empregada a técnica estatística de validação cruzada (Figura 20), que divide o conjunto de dados em  $k$  subconjuntos (ou *folds*) treinando o modelo  $k$  vezes, utilizando  $k-1$  subconjuntos para treinamento e o subconjunto restante para teste, com retorno dos resultados para cada rodada de treinamento/teste.

**Figura 22:** Divisão dos Dados de Entrada em conjuntos de Treino/Teste para realização da validação cruzada e Conjunto de Validação para testar a capacidade de generalização do modelo



Fonte: Adaptado de *scikit-learn* (2024)

## 4.6 Configurações

### 4.6.1 Configuração dos Modelos Machine Learning

Para avaliar o desempenho dos algoritmos, as melhores arquiteturas foram definidas por meio da utilização da função *GridSearchCV* da biblioteca *scikit-learn* no Python 3.10, que analisa diferentes combinações de parâmetros pré-definidos, analisando o desempenho do algoritmo testado em cada combinação, e ao fim, escolhe os parâmetros que tiveram os melhores resultados utilizando validação cruzada. Para comparar dois modelos diferentes, neste caso RF e XGB, é preciso conduzir dois *GridSearchCV*, pois cada um é configurado com os parâmetros

específicos de cada modelo (Quadro 4) e aplicado aos dados de treino e teste para encontrar a melhor combinação.

**Quadro 5:** Variações de parâmetros testadas para os modelos Random Forest e XGBoost utilizando GridSearchCV

ALGORITMO	PARÂMETRO	CÓDIGO	VARIAÇÕES
RANDOM FOREST	Número Máximo de Árvores	<i>n_estimators</i>	<ul style="list-style-type: none"> <li>• 100 árvores</li> <li>• 150 árvores</li> <li>• 200 árvores</li> <li>• 250 árvores</li> </ul>
	Profundidade Máxima	<i>max_depth</i>	<ul style="list-style-type: none"> <li>• None</li> <li>• 10</li> <li>• 20</li> <li>• 30</li> </ul>
XGBOOST	Número Máximo de Árvores	<i>n_estimators</i>	<ul style="list-style-type: none"> <li>• 100 árvores</li> <li>• 150 árvores</li> <li>• 200 árvores</li> <li>• 250 árvores</li> </ul>
	Taxa de Aprendizagem	<i>learning_rate</i>	<ul style="list-style-type: none"> <li>• 0,001</li> <li>• 0,01</li> <li>• 0,1</li> </ul>
	Profundidade Máxima	<i>max_depth</i>	<ul style="list-style-type: none"> <li>• None</li> <li>• 10</li> <li>• 20</li> <li>• 30</li> </ul>

**Fonte:** Elaborado pelo Autor (2024)

Para o modelo RF, foi utilizada a classe *RandomForestRegressor* do módulo *ensemble*, pertencente a biblioteca *scikit-learn*. Os hiperparâmetros explorados foram o número máximo de árvores na floresta (parâmetro *n\_estimators*) variando entre 100, 150, 200 e 250 árvores, e também, a profundidade máxima de cada árvore de decisão da floresta (parâmetro *max\_depth*), que variou entre 10, 20, 30 e sem limites. Já para o modelo XGB, foi utilizada a classe *XGBRegressor* da biblioteca *xgboost*, sendo testados os mesmos parâmetros do modelo RF, além da variação da taxa de aprendizagem (parâmetro *learning\_rate*) para 0,001, 0,01 e 0,1.

#### 4.6.2 Configuração de Hardware e Software

Para a realização das etapas de treinamento e validação, foi utilizado um computador com as seguintes configurações: processador Intel® Core™ i5 – 7400 e 12GB de memória RAM. Com relação as configurações de software, foi utilizada a

linguagem Python versão 3.10, com o Spyder versão 5.4.3 como ambiente de desenvolvimento (*Integrated Development Environment* – IDE) e para elaboração dos mapas foi utilizado o software QGIS versão 3.28.15.

## 4.7 Avaliação de Desempenho

### 4.7.1 Métricas de Desempenho

Para verificar o desempenho de cada um dos modelos em prever os valores da variável resposta foram avaliadas três métricas, sendo elas: o *Mean Absolute Error* – MAE, o *Root Mean Squared Error* – RMSE e o Coeficiente de Determinação –  $R^2$ , dadas por:

$$MAE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (7)$$

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (8)$$

$$R^2(y, \hat{y}) = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (9)$$

Onde,  $y$  representa os valores reais observados,  $\hat{y}$  representa os valores preditos e  $\bar{y}$  representa a média dos valores de  $y$ . Os resultados de  $R^2$  indicam que quanto mais próximo de 1, mais explicativo é o modelo em relação aos dados previstos. Diferente do  $R^2$ , os valores do MAE não possuem um limite máximo e medem a diferença entre os valores reais e os valores preditos, enquanto os valores do RMSE, apesar de terem a mesma escala dos valores de MAE, acabam penalizando possíveis valores discrepantes, os chamados *outliers* (Harrison, 2020). Para a métrica  $R^2$  os resultados mais próximos de um são considerados “melhores”, enquanto que para as métricas de erro (MAE e RMSE), quanto menores forem os valores, melhores serão os resultados.

Desse modo, as métricas são calculadas na validação cruzada utilizando a função *cross\_validate* da biblioteca *scikit-learn*, que permite a determinação do número de *folds* (neste caso,  $k = 10$ ) e das métricas que serão avaliadas (MAE, RMSE e  $R^2$ ). Em seguida, foi aplicado o teste não paramétrico de Mann-Whitney para comparar os dois modelos de algoritmos e determinar o melhor desempenho.

Por fim, o modelo foi avaliado com os dados de validação separados anteriormente, testando seu desempenho com dados que foram deixados de fora da fase de treinamento/teste.

#### 4.7.2 Análises Estatísticas

A correlação de Spearman é uma medida não paramétrica acerca da relação entre duas variáveis, podendo variar entre  $-1$  e  $+1$ . Quanto mais próxima dos extremos, mais forte a relação entre as variáveis, enquanto que valores mais próximos de zero indicam correlações fracas (Spearman, 1904). A correlação de Spearman foi aplicada utilizando a função *spearmanr* pertencente ao módulo *stats* da biblioteca *SciPy* do Python.

O teste não paramétrico de Mann-Whitney U é usado para determinar se existe diferença estatisticamente significativa entre as médias de duas amostras independentes (Mann e Whitney, 1947). O teste de Mann-Whitney é adequado neste caso por conta de o tamanho das amostras serem pequenos (10 valores) e pela suspeita de que as distribuições das amostras não são seguem uma distribuição normal. Neste teste, a hipótese nula ( $H_0$ ) aponta que não há diferença significativa entre as médias e que os dois grupos são amostrados de populações com distribuições idênticas, enquanto que a hipótese alternativa ( $H_1$ ) aponta que há diferenças entre os dois grupos.

Dessa forma, o teste de Mann-Whitney foi aplicado utilizando linguagem Python, por meio da função *scipy.stats.mannwhitneyu*, presente no módulo *stats* da biblioteca *SciPy*, aos resultados obtidos de  $R^2$ , RMSE e MAE. Foi adotado ainda um nível de significância de 5%, ou seja, o valor-p resultante dos testes precisa ser menor que 0,05 para que a hipótese nula possa ser rejeitada.

O teste não-paramétrico dos postos sinalizados de Wilcoxon (1970) é usado para comparar duas amostras pareadas, mas que também pode ser aplicado a uma amostra em relação a um valor de referência (Neuhauser, 2011). O teste é aplicado sobre a diferença entre os valores de uma amostra pareada e um valor de referência  $\mu$ , de forma que a hipótese nula ( $H_0$ ) aponta que não há diferença significativa entre a amostra e o valor de referência ( $p > 0,05$ ), enquanto a hipótese alternativa ( $H_1$ ) afirma que a mediana das diferenças é diferente de zero ( $p < 0,05$ ).

Por sua vez, o teste de Wilcoxon foi aplicado por meio da função `scipy.stats.wilcoxon`, presente no módulo `stats` da biblioteca `SciPy`, aos resultados obtidos de  $R^2$  e de RMSE. Foi adotado um nível de significância de 5% para ambos os testes.

#### **4.8 Testes Preliminares – *Benchmark***

Com o intuito de comparar e verificar a aplicabilidade das etapas metodológicas sugeridas, bem como provar os conceitos e a metodologia apresentados, testando o conhecimento das técnicas envolvidas, foi proposta a realização de um teste de *benchmark* entre os modelos apresentados e um modelo com desempenho já conhecido e validado, com o objetivo de verificar a viabilidade deste projeto, estabelecendo uma referência para as etapas futuras. Para este fim, 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). O caso foi escolhido por se tratar de uma metodologia semelhante de aquisição e preparo dos dados, além de tratarem do mesmo assunto – Incêndios Florestais, mesmo que em contextos distintos. O referido trabalho teve por objetivo avaliar a predição de fogo em escala regional e fornecer saídas compreensíveis, na forma de regras de decisão, capazes de predizer os valores de risco médio para cada célula de grade, definindo assim as unidades de manejo do fogo no estado do Espírito Santo.

A planilha de dados foi fornecida diretamente pelo autor da pesquisa, contendo uma variável-alvo, representada pela Densidade Kernel calculada sobre 3.314 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.

Dessa forma, os dados foram inseridos diretamente em ambiente de programação Python, aonde foram pré-processados e tratados. Posteriormente, o *dataset* com os dados foi submetido ao algoritmo RF, a partir de validação cruzada utilizando a técnica *K-fold* com  $k = 10$ , onde o conjunto de dados é dividido em um número  $k$  de blocos (*folds*), realizando  $k$  iterações utilizando diferentes dados em cada uma e, portanto, tendo resultados diferentes em cada iteração do modelo (Witten e Frank, 2002). Finalmente as métricas resultantes foram comparadas às obtidas por Juvanhol *et al.* (2023) ao utilizar o algoritmo de Árvore de Classificação e Regressão (CART – *Classification And Regression Tree*).

Por fim, na busca por determinar qual dos modelos teve um melhor desempenho em relação a cada uma das métricas, considerando que os dados apresentados para o CART não possuem repetições, foi aplicado o teste não-paramétrico dos postos sinalizados de Wilcoxon (1970).

## **4.9 Predição da Susceptibilidade a Incêndios**

### **4.9.1 Mapa de Predição da Susceptibilidade de Incêndios Florestais**

A partir da escolha do modelo mais adequado através dos testes descritos anteriormente, o modelo foi salvo e aplicado ao conjunto total de dados, gerando os valores de predição. Esses valores foram carregados em ambiente SIG e classificados em cinco classes de susceptibilidade a incêndios: muito baixa, baixa, média, alta e muito alta. O método estatístico utilizado para a classificação foi o algoritmo de quebras naturais de Jenks que ajusta as quebras dos intervalos de classe, de modo a minimizar a variação interna dentro de cada classe e maximizar a variação entre as classes (Jenks e Caspall, 1971).

### **4.9.2 Importância e Contribuição das Variáveis**

É possível analisar os modelos RF e XGB de modo a compreender a importância de cada variável preditora. A importância de cada uma das variáveis utilizadas na etapa de treinamento foi obtida através da função *feature\_importances\_* em ambos os modelos. Outra técnica considerada nessa análise foi a contribuição de cada variável para as previsões realizadas. Neste caso, foi utilizada a abordagem

baseada nos valores SHAP (*Shapley Additive Explanations*), que atribui uma contribuição específica de cada variável para cada previsão do modelo. Ou seja, diferente do *feature importance* que identifica as variáveis mais influentes na capacidade preditiva global do modelo, a *feature contribution* possibilita a compreensão de como cada variável contribui para as previsões do modelo em nível individual, por meio da contribuição média ao longo de todas as previsões (Marcílio e Eler, 2020).

Os valores de *feature contribution* apontam a influência entre cada variável e os resultados preditos, essa influência pode ser positiva ou negativa, ou seja, se à medida que a variável aumenta, o resultado também tende a aumentar, a contribuição será positiva; ou se à medida que a variável aumenta, o resultado diminui, a contribuição será negativa.

## 5 RESULTADOS E DISCUSSÃO

### 5.1 Testes Preliminares – Benchmark

Os resultados dos testes preliminares foram apresentados no XXVI Encontro de Modelagem Computacional, por meio do artigo “Avaliação Exploratória do Algoritmo RF para Previsão de Incêndios Florestais”. A informação sobre autoria e o resumo do artigo estão no Apêndice A.

A Tabela 1 apresenta os resultados das 10 execuções da validação cruzada para o RF, além dos valores estatísticos de mínimo, máximo, mediana, média e desvio-padrão, e comparados com os resultados obtidos para o algoritmo CART no trabalho de Juvanhol *et al.* (2023). O teste não paramétrico de Wilcoxon apontou que há diferença estatisticamente significativa a um nível de significância de 5% entre os valores de RF e CART para ambas as métricas avaliadas neste estudo.

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 RF. Por outro lado, o algoritmo CART apresentou um Erro Quadrático Médio (MSE – *Mean Squared Error*) médio de 0,01 com uma configuração de árvore com 5.137 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 dos resultados obtidos a partir da validação cruzada (k = 10) para o algoritmo RF com o resultado médio obtido por Juvanhol *et al.* (2023) para o Algoritmo CART utilizando o mesmo conjunto de dados

ALGORITMO		R <sup>2</sup>	RMSE
RANDOM FOREST	VALIDAÇÃO CRUZADA (K = 10)		
	SPLIT 1	0,9877	0,0106
	SPLIT 2	0,9884	0,0103
	SPLIT 3	0,9874	0,0108
	SPLIT 4	0,9891	0,0099
	SPLIT 5	0,9890	0,0101
	SPLIT 6	0,9880	0,0104
	SPLIT 7	0,9891	0,0100
	SPLIT 8	0,9885	0,0103
	SPLIT 9	0,9884	0,0103
	SPLIT 10	0,9885	0,0103
MÍNIMO	0,9874	0,0099	
MÁXIMO	0,9891	0,0108	
MEDIANA	0,9884	0,0103	
DESVIO PADRÃO	0,0002	0,0005	
<b>MÉDIA</b>	0,9884	0,0103	
<b>*CART (MÉDIA)</b>		0,88	0,55

\* Resultado obtido por Juvanhol *et al.* (2023)

Fonte: Elaborado pelo Autor (2024)

Ao comparar as pontuações de influência de cada variável nos modelos CART e RF (Quadro 5) as maiores variações foram do Uso e Cobertura da Terra (1,89% a 75,07%), Índice Topográfico Composto (3,79% a 24,06%), Proximidade de Estradas (2,14% a 9,92%) e Declividade (0,94% a 9,14%). 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.

**Quadro 6:** Comparação em termos percentuais da Importância de cada variável para os Modelos RF e CART

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

\* Resultado obtido por Juvanhol *et al.* (2023)

**Fonte:** Elaborado pelo Autor (2024)

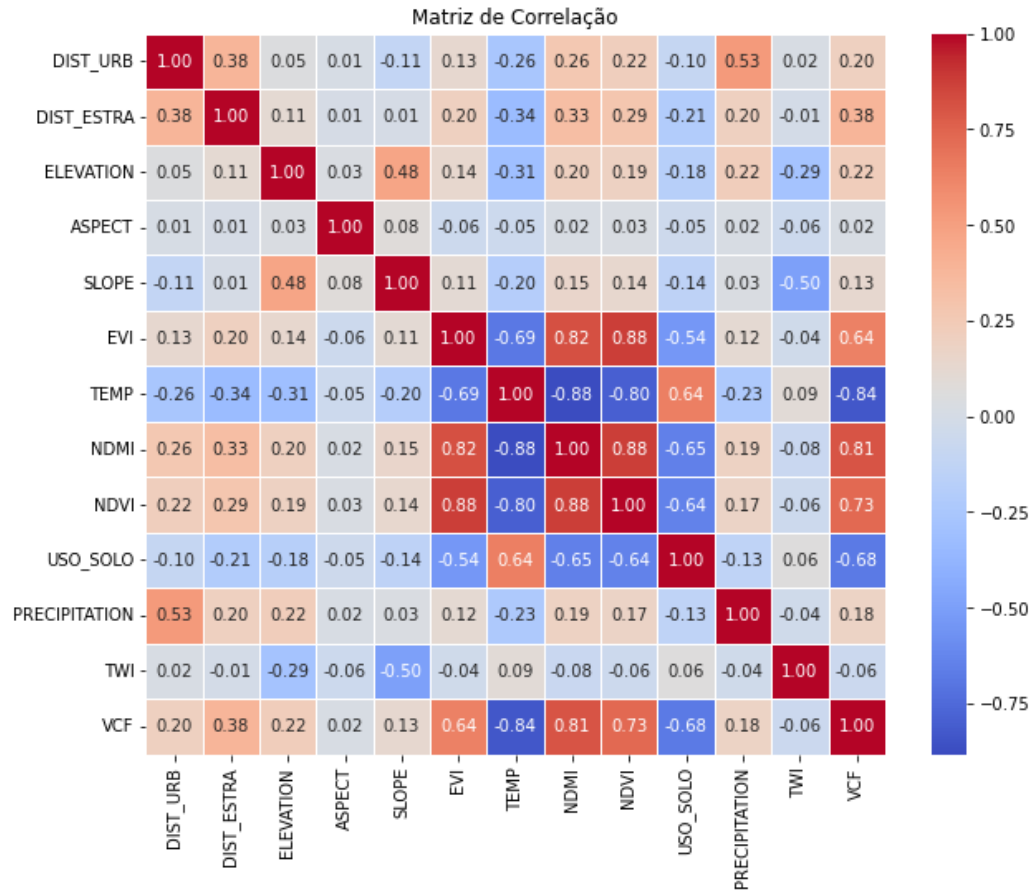
## 5.2 Estudo de Caso – APA Triunfo do Xingu

### 5.2.1 Seleção de Variáveis

A partir do cálculo das correlações, foi elaborada a matriz de correlação entre as 13 variáveis consideradas, apresentada na figura 21.

Entre as relações com alta correlação entre si, temos o EVI e o NDVI com 0,88. Para essa relação, o NDVI foi mantido por ser mais amplamente utilizado em diversas abordagens, em relação à predição de incêndios florestais, em comparação com o EVI. Por outro lado, as demais relações com alta correlação entre si demandaram uma análise mais aprofundada. As variáveis NDMI e VCF apresentaram alta correlação entre si (0,81) e também com as variáveis Temperatura (-0,88 e -0,84, respectivamente) e NDVI (0,88 e 0,73, respectivamente). Da mesma forma que o NDVI, a temperatura foi mantida por conta da sua relação direta com a ocorrência e a detecção de incêndios, restando a escolha entre NDMI e VCF.

**Figura 23:** Matriz de Correlação Entre as variáveis consideradas para serem predictoras de Incêndios Florestais



**Fonte:** Elaborado pelo Autor (2024)

Neste estudo, a preferência por manter o VCF se deu por conta do NDMI, que representa a umidade da vegetação, apresentar informações que podem ser contempladas em outras variáveis, como a precipitação e a temperatura, podendo ser inferidas de maneira indireta. Outro fator importante é o fato de o NDMI ser mais relevante em regiões onde a vegetação esteja sujeita a situações de estresse hídrico (Rahman e Mesev, 2019), ou seja, em locais onde há uma variação extrema no volume de chuvas, ao contrário do que acontece na região amazônica, onde apesar de haver períodos de maior e menor precipitação, não há grandes variações entre os períodos.

### 5.2.2 Parâmetros dos Modelos Machine Learning

Os resultados de cada uma das variações testadas, bem como os tempos médios de treinamento, para os algoritmos RF e XGB estão disponíveis no Apêndice

B. Dessa forma, a melhor configuração encontrada para o algoritmo RF (Quadro 6) consistiu em 200 árvores ( $n\_estimators = 200$ ) e sem profundidade máxima ( $max\_depth = None$ ). Enquanto que para o algoritmo XGB, a configuração que apresentou o melhor desempenho (Quadro 6) foi a que consistia de 250 árvores ( $n\_estimators = 250$ ), profundidade máxima de 20 ( $max\_depth = 20$ ) e uma taxa de aprendizagem de 0,1 ( $learning\_rate = 0,1$ ).

**Quadro 7:** Melhores Combinações de parâmetros encontradas para os modelos RF e XGB utilizando GridSearchCV

ALGORITMO	PARÂMETRO	MELHOR COMBINAÇÃO
RANDOM FOREST	Número Máximo de Árvores <i>n_estimators</i>	200 Árvores
	Profundidade Máxima <i>max_depth</i>	None
XGBOOST	Número Máximo de Árvores <i>n_estimators</i>	250 árvores
	Taxa de Aprendizagem <i>learning_rate</i>	0.1
	Profundidade Máxima <i>max_depth</i>	20

**Fonte:** Elaborado pelo Autor (2024)

### 5.2.3 Desempenho dos Algoritmos

Os resultados das 10 execuções da validação cruzada dos algoritmos RF e XGB avaliados pelas métricas RMSE, MAE e  $R^2$  dos dados de treino estão detalhados na Tabela 2 e representados na Figura 22.

Ao realizar o teste não-paramétrico de Mann-Whitney para as métricas RMSE e  $R^2$ , não houve diferença estatisticamente significativa entre os dois algoritmos, a um nível de confiança de 95%. Contudo, ao comparar os resultados de MAE, foi possível verificar diferença estatisticamente significativa.

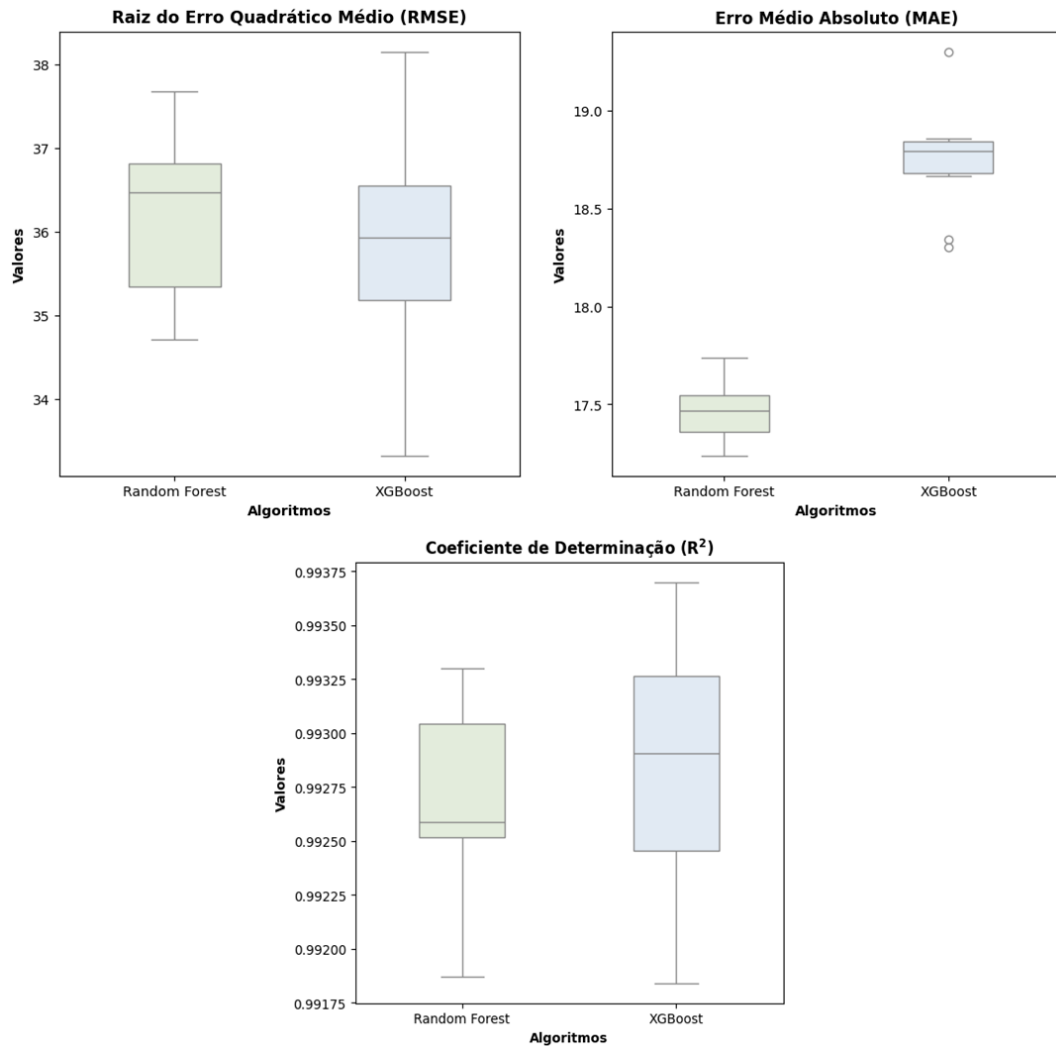
**Tabela 2:** Valores da Validação Cruzada, Média, Desvio Padrão, Mediana, Mínimos e Máximos de Raiz do RMSE, MAE e  $R^2$  para os algoritmos RF e XGB usando os dados de teste/treino

RANDOM FOREST				XGBOOST				
Tempo de Execução: 3947s				Tempo de Execução: 1141s				
	RMSE	MAE	$R^2$		RMSE	MAE	$R^2$	
VALIDAÇÃO CRUZADA (K = 10)	SPLIT 1	36,8718	17,4682	0,9925	SPLIT 1	35,8865	18,8054	0,9929
	SPLIT 2	34,8124	17,2537	0,9933	SPLIT 2	33,5265	18,6657	0,9936
	SPLIT 3	34,7051	17,2384	0,9932	SPLIT 3	35,0228	18,3015	0,9931
	SPLIT 4	36,2188	17,4631	0,9926	SPLIT 4	34,7056	18,7247	0,9932
	SPLIT 5	36,6295	17,5724	0,9926	SPLIT 5	38,4170	19,2998	0,9918
	SPLIT 6	36,6142	17,5465	0,9925	SPLIT 6	37,1396	18,8565	0,9923
	SPLIT 7	37,6799	17,5411	0,9918	SPLIT 7	36,0321	18,7801	0,9927
	SPLIT 8	36,3186	17,4276	0,9925	SPLIT 8	36,9110	18,8405	0,9922
	SPLIT 9	37,0203	17,7360	0,9924	SPLIT 9	35,9579	18,8380	0,9928
	SPLIT 10	35,0535	17,3349	0,9931	SPLIT 10	33,7524	18,3432	0,9936
<b>MÉDIA</b>	36,2616	17,4582	0,9926	<b>MÉDIA</b>	35,7351	18,7455	0,9929	
<b>DESVIO PADRÃO</b>	1,0313	0,1453	0,0004	<b>DESVIO PADRÃO</b>	1,4520	0,2660	0,0006	
<b>MEDIANA</b>	36,4990	17,4656	0,9926	<b>MEDIANA</b>	35,9222	18,7927	0,9929	
<b>MÍNIMO</b>	34,5729	17,2384	0,9918	<b>MÍNIMO</b>	33,5265	18,3015	0,9918	
<b>MÁXIMO</b>	38,0857	17,7360	0,9933	<b>MÁXIMO</b>	38,4170	19,2998	0,9936	

Fonte: Elaborado pelo Autor (2024)

Deste modo, o desempenho de ambos os modelos se mostrou satisfatório, e então, optou-se por utilizar apenas o algoritmo XGB, que possui um menor custo computacional, representado pelo tempo de execução na Tabela 2, além da escalabilidade para possível aumento na quantidade de dados, entre outras vantagens (Chen e Guestrin, 2016). Além disso, devido à similaridade dos desempenhos, não houve diferenças perceptíveis no estudo do mapa de predição de susceptibilidade de incêndios.

**Figura 24:** Boxplots dos valores de validação cruzada para RMSE, MAE e  $R^2$  para os algoritmos RF e XGB usando os dados de teste/treino



**Fonte:** Elaborado pelo Autor (2024)

A utilização de algoritmos ensemble tem papel de destaque no que se refere a predição de incêndios florestais. Em estudos que utilizaram a abordagem de regressão, como o de Xie e Peng (2019), Shmuel e Heifetz (2023); quanto em trabalhos que utilizaram a abordagem da classificação, como Michael *et al.* (2021), Bilucan *et al.* (2022), Seddouki *et al.* (2023) e Agrawal *et al.* (2023), os algoritmos *ensemble*, RF e XGB, tiveram desempenhos superiores aos demais modelos testados. Esses trabalhos também corroboram a proximidade do desempenho entre os dois algoritmos, contudo com o algoritmo XGB apresentando melhor performance nos valores unitários nos testes empregados.

### 5.2.4 Resultados dos Dados de Validação

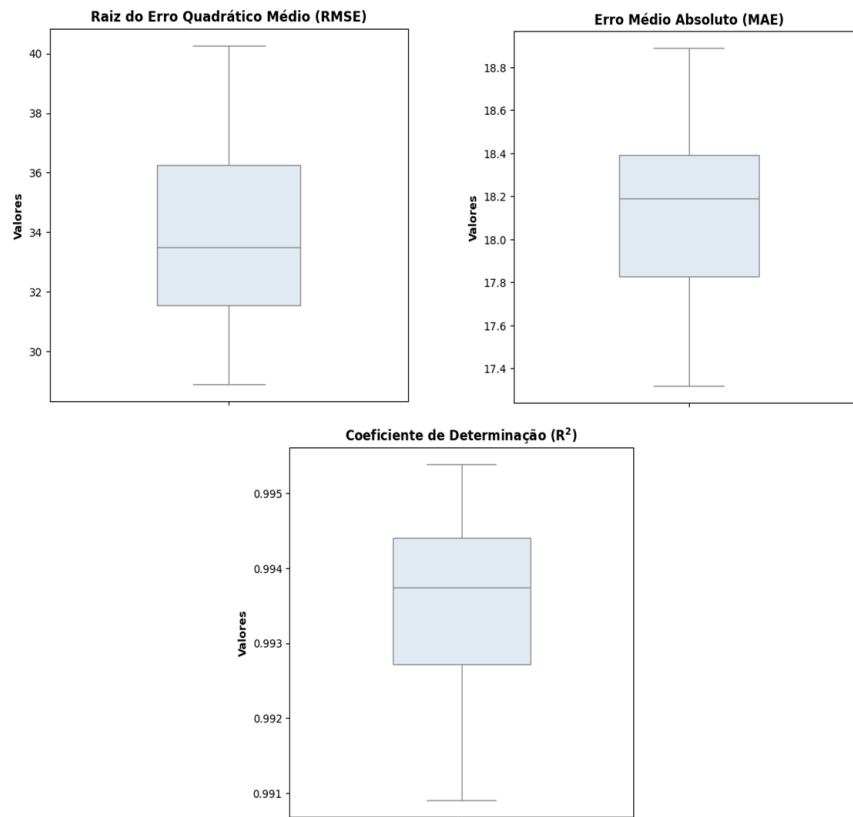
A capacidade de generalização do modelo XGB foi testada aplicando-se o modelo gerado a partir da etapa de treino/teste, a um conjunto de dados que foi deixado de fora da etapa de treino\teste, conforme ilustrado na figura 20. A Tabela 3 e a Figura 23 apresentam os resultados obtidos na validação do modelo, RMSE médio de 33,9902, um MAE médio de 18,1331 e um  $R^2$  médio de 0,9934, mostrando que o algoritmo foi capaz de realizar a predição de forma confiável, obtendo resultados de validação condizentes com os resultados obtidos na fase de treino/teste.

**Tabela 3:** Valores da Validação Cruzada, Média, Desvio Padrão, Mediana, Mínimos e Máximos de RMSE, MAE e  $R^2$  para o algoritmo XGB usando os dados de validação

		XGBOOST		
		RMSE	MAE	$R^2$
<b>CROSS-VALIDATION (K = 10)</b>	SPLIT 1	36,0005	18,0384	0,9927
	SPLIT 2	36,3482	18,4150	0,9927
	SPLIT 3	31,4825	18,3092	0,9944
	SPLIT 4	40,2581	18,8889	0,9909
	SPLIT 5	28,8914	17,3185	0,9953
	SPLIT 6	33,6949	17,6668	0,9935
	SPLIT 7	33,3125	18,3291	0,9939
	SPLIT 8	31,7255	18,0739	0,9943
	SPLIT 9	31,3502	17,7564	0,9945
	SPLIT 10	36,8385	18,5347	0,9923
	<b>MEAN</b>	33,9902	18,1331	0,9934
	<b>STD. DV.</b>	3,1926	0,4378	0,0012
	<b>MEDIAN</b>	33,5037	18,1331	0,9937
	<b>MIN.</b>	28,8914	17,3185	0,9909
	<b>MAX.</b>	40,2581	18,8889	0,9953

**Fonte:** Elaborado pelo Autor (2024)

**Figura 25:** Boxplots dos valores de validação cruzada para RMSE, MAE e  $R^2$  para o algoritmo XGB usando os dados de validação

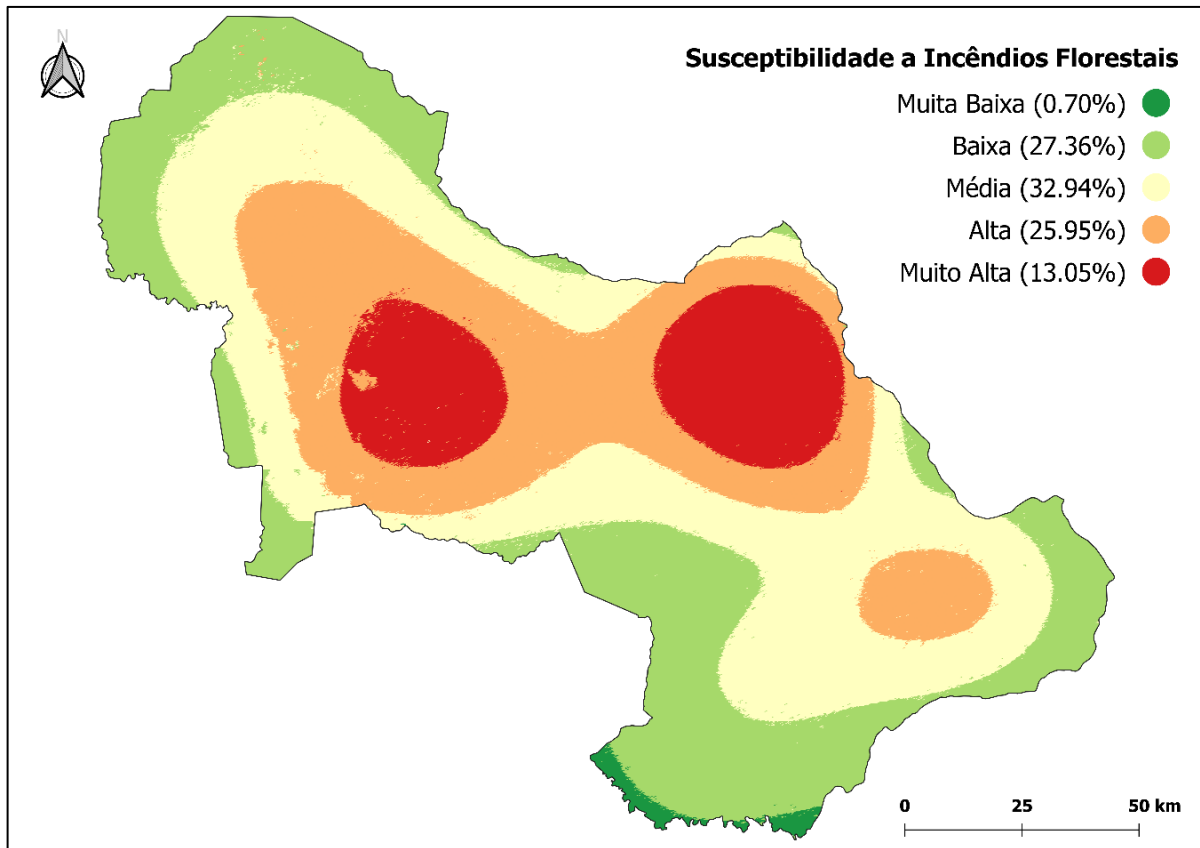


**Fonte:** Elaborado pelo Autor (2024)

### 5.2.5 Mapa de Predição da Susceptibilidade de Incêndios Florestais

A partir do modelo validado, o mapa de susceptibilidade de incêndios florestais foi gerado a partir da classificação dos valores de densidade de queima em muito baixo, baixo, médio, alto e muito alto (Figura 24). As áreas de susceptibilidade a incêndios florestais muito baixa, em verde escuro, representam apenas 0,7% da área total da APA e, estão localizadas nas extremidades sul, centro-oeste e norte. Por outro lado, as áreas de susceptibilidade baixa, em verde claro, e média, em amarelo, representam 27,36% e 32,94% da área total, respectivamente, representando o panorama geral da APA, com áreas irradiando das bordas em direção ao centro. As áreas de alta susceptibilidade se concentram nas regiões centrais a leste e a oeste e também no sudeste da APA, representando 25,95% da área total, assim como as regiões que requerem um nível ainda maior de atenção, com susceptibilidade a incêndios florestais muito alta, ocupando 13,05% da extensão total da área, que também se concentram na porção central leste e também oeste.

**Figura 26:** Mapa de Predição da Susceptibilidade a Incêndios Florestais na APA Triunfo do Xingu, Pará, Brasil



**Fonte:** Elaborado pelo Autor (2024)

Por meio do Quadro 7, as áreas com susceptibilidade muito alta podem ser comparadas às áreas com muito baixa e baixa susceptibilidade em relação a cada uma das variáveis preditoras. Em relação à distância das estradas, as regiões de menor susceptibilidade estão a uma distância média de 3 km de alguma estrada, enquanto as áreas de maior susceptibilidade estão distantes 2 km de qualquer acesso. Já com relação a áreas habitadas, áreas de baixa e muito baixa susceptibilidade estão um pouco mais distantes (40 km, em média), se comparadas às áreas de maior susceptibilidade (38 km, em média).

Em relação às características topográficas, ambas as regiões possuem um perfil muito similar entre si e a outras classes de susceptibilidade no território da APA. São regiões de baixa elevação, em média 260 metros acima do nível do mar, com baixos valores de inclinação, entre 5 e 7%, pouco acúmulo de água na superfície e com distribuição de suas vertentes, com leve predominância voltadas para o norte.

Por outro lado, analisando as variáveis ambientais, há algumas diferenças notáveis, principalmente em relação à temperatura média, precipitação média anual, os tipos de uso do solo e a porcentagem de cobertura vegetal. As áreas de baixa e muito baixa susceptibilidade possuem uma temperatura média 1,5 °C mais baixa em relação às áreas de maior susceptibilidade, 32 °C e 33,5 °C, respectivamente; da mesma forma, em relação à precipitação média anual, áreas mais susceptíveis a incêndios apresentam, em média, 2000 mm.ano<sup>-1</sup>, enquanto as áreas onde há menor probabilidade de incêndios têm, em média, 2200 mm.ano<sup>-1</sup>.

Outro fator de destaque são as classes de uso e cobertura predominantes em cada uma dessas áreas, uma vez que, enquanto há uma predominância de áreas de floresta nas áreas de menor susceptibilidade, nas áreas com muito alta susceptibilidade predominam áreas de pastagem. Por fim, em relação à vegetação, tanto as áreas menos susceptíveis quanto as mais susceptíveis a incêndios possuem vegetações, em média, saudáveis, apresentando valores de NDVI de, em média, 0,8, contudo, ao observar os valores de VCF, as áreas com baixa susceptibilidade possuem uma maior cobertura vegetal, com uma média de 56% em relação às áreas de maior susceptibilidade a incêndios (47%).

A predominância de áreas de pastagem nas regiões mais susceptíveis a incêndios passa diretamente pelo processo de transformação dos ecossistemas naturais em agroecossistemas de produção, no qual, o fogo é utilizado na queima de florestas e no manejo das pastagens (Vasconcelos *et al.*, 2013). Esse processo é apontado pelo Instituto do Homem e Meio Ambiente da Amazônia (IMAZON), como estando relacionado ao desflorestamento ocasionado pela exploração madeireira e o crescimento da pecuária (Souza JR *et al.*, 2018). Os trabalhos de Dos Santos *et al.* (2020, 2023), realizados na região da APA Triunfo do Xingu e no município de São Félix do Xingu, identificaram que nos últimos 10 anos, 12 mil km<sup>2</sup> de floresta foram substituídos por pastagem nessas regiões, onde as práticas de manejo empregadas tornam-se fontes dispersoras de fogo para as áreas de floresta.

**Quadro 8:** Caracterização das Áreas de Muito Baixa e Baixa Susceptibilidade em relação às Áreas de Muito Alta Susceptibilidade a incêndios florestais considerando as variáveis preditoras estudadas

VARIÁVEL		ÁREAS DE SUSCEPTIBILIDADE BAIXA E MUITO BAIXA	ÁREAS DE SUSCEPTIBILIDADE MUITO ALTA
<b>SOCIOECONÔMICAS</b>	Distância de Estradas	Média = 3 km de distância	Média = 2 km de distância
	Distância de Áreas Habitadas	Média = 40 km de distância	Média = 38 km de distância
<b>TOPOGRÁFICAS</b>	Elevação	Média = 269 metros	Média = 253 metros
	Declividade	Média = 7,3%	Média = 5,2%
	TWI	Média = 10,45	Média = 10,6
	Orientação de Vertentes	Norte (26,52%) Leste (25,01%) Sul (24,41%) Oeste (24,04%)	Norte (25,71%) Leste (25,83%) Sul (24,11%) Oeste (24,33%)
<b>AMBIENTAIS</b>	Temperatura	Média = 32 °C	Média = 33,5 °C
	Precipitação	Média = 2200 mm.ano <sup>-1</sup>	Média = 2000 mm.ano <sup>-1</sup>
	Uso e Cobertura da Terra	Floresta (67,57%) Pastagem (31,24%) Outros (1,2%)	Pastagem (53,58%) Floresta (43,41%) Outros (3%)
	NDVI	Média = 0,8	Média = 0,8
	VCF	Média = 56%	Média = 47 %

Fonte: Elaborado pelo Autor (2024)

A pouca variação dos valores de NDVI entre as áreas de menor susceptibilidade e de maior susceptibilidade reflete a baixa influência de fatores relacionados à vegetação sobre a ocorrência de incêndios. Os valores de NDVI apresentados, em torno de 0,8 para ambas as áreas, sugerem vegetações saudáveis e densas. Ao mesmo tempo, as áreas menos susceptíveis são observadas estritamente em locais com maior continuidade da vegetação florestal (56%), o que acaba por formar uma densa barreira natural contra a propagação de incêndios (Loudermilk *et al.*, 2022, Tian *et al.*, 2022). Por outro lado, mesmo que os valores de NDVI nas áreas mais susceptíveis sejam altos, a predominância de áreas de pastagem sugere uma vegetação mais inflamável por conta de sua estrutura simples e uniforme (Neary e Leonard, 2020; Yan e Liu, 2021), além de ser uma vegetação mais fragmentada, como apontam os valores de VCF, facilitando a ignição e propagação do fogo.

#### 5.2.6 Influência e Contribuição das Variáveis

O Quadro 8 apresenta a importância percentual relativa de cada variável independente na explicação da variabilidade da variável resposta, neste caso os valores preditos de densidade de queima, de forma que quanto maior o valor de importância, mais impacto aquela variável tem na capacidade preditiva do modelo (Casalicchio *et al.* 2019).

A variável precipitação apresentou a maior importância relativa percentual para o modelo, com 70.45%, sugerindo que precipitação anual média é um fator significativo na susceptibilidade de ocorrência de queimadas. Em seguida, a distância de áreas urbanas, com 15.02%, e o uso do solo, com 7.87%, também se apresentam como variáveis importantes. Outras variáveis como Elevação, Distância de Estradas e Temperatura têm influências percentuais menores, mas ainda desempenham papéis importantes na susceptibilidade a queimadas, embora em menor grau em comparação às anteriores. As demais variáveis apresentam influências muito baixas, sugerindo que elas têm um impacto mínimo no modelo, contudo é importante analisar o contexto geral e não apenas cada variável isoladamente, pois mesmo uma variável tendo uma influência muito baixa, ela ainda pode adicionar informações importantes ao modelo. Ou seja, mesmo uma variável apresente pouca importância para redução das

incertezas nas previsões do modelo, esta é relevante para a ocorrência do incêndio, pois a relação causal ou linear entre as variáveis não é mensurada pelo parâmetro *feature importance*.

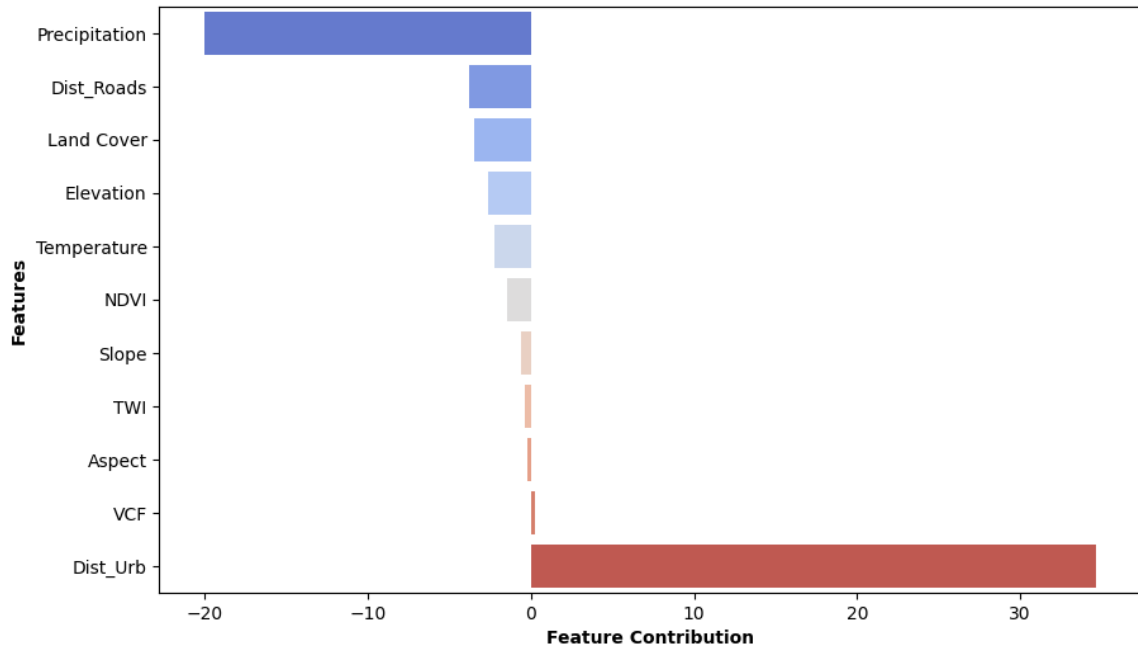
**Quadro 9:** Importância Percentual de cada variável preditora para o modelo de predição XGBoost

VARIÁVEL	IMPORTÂNCIA
Precipitação	70,45%
Distância de Áreas Habitadas	15,02%
Uso e Cobertura da Terra	7,87%
Elevação	2,29%
Temperatura	1,79%
Distância de Estradas	1,38%
VCF	0,62%
NDVI	0,41%
Declividade	0,07%
TWI	0,03%
Orientação de Vertentes	0,02%

**Fonte:** Elaborado pelo Autor (2024)

Por outro lado, a Figura 25 apresenta a contribuição média de cada variável independente para a previsão do valor resposta. A variável com maior contribuição absoluta média individual foi a Distância para áreas urbanas, com uma contribuição positiva de 34,66, seguida pela precipitação, com uma contribuição negativa de -20,06, seguidas por Distância de Estradas (-3,82), Uso do Solo (-3,46), Elevação (-2,60), Temperatura (-2,26) e NDVI (-1,48), enquanto Orientação de Vertentes, Declividade, TWI e VCF não tiveram contribuições significativas.

**Figura 27:** Contribuição Média de cada variável preditora para as previsões realizadas pelo modelo XGB de Predição de Susceptibilidade a Incêndios Florestais



**Fonte:** Elaborado pelo Autor (2024)

A alta influência da precipitação, aliada à sua contribuição negativa, indicam que a precipitação média anual tem um papel de destaque na predição da susceptibilidade de incêndios florestais, diminuindo os valores de ocorrência de queimas nas áreas de maior precipitação. Pourghasemi *et al.* (2020b), Mohajane *et al.* (2021), Celis *et al.* (2023), Juvanhol *et al.* (2023) e Hang *et al.* (2024) ao estudarem a ocorrência e a predição de incêndios utilizando técnicas de ML, também apontaram a precipitação como um dos fatores preditores mais importantes, destacando a influência das chuvas na condição do material combustível, tornando-o menos inflamável em períodos mais chuvosos.

Ao contrário dos estudos citados anteriormente, realizados em outros lugares do mundo, em que a precipitação, com uma maior variação ao longo do ano e um volume acumulado menor, potencializa a ocorrência de incêndios, na APA Triunfo do Xingu esta variável está relacionada a contribuir como limitante da propagação dos incêndios e a mitigação de seus impactos. Essa relação é corroborada por Dos Santos *et al.* (2023), ao identificarem que menores volumes de precipitação não estão entre os fatores determinantes para ocorrência de incêndios na Amazônia, por conta do elevado volume de precipitação acumulada anual, o que permite que esse recurso esteja disponível regularmente.

A Distância de áreas urbanas, apesar de ter uma importância percentual menor que a precipitação (Quadro 8), apresentou uma contribuição absoluta média alta (Fig.10), o que evidencia sua relevância no estudo da predição de incêndios. Trabalhos anteriores, como os de Bui *et al.* (2017), Kalantar *et al.* (2020) e Pham *et al.* (2020), ao utilizarem diferentes algoritmos de ML para prever a susceptibilidade de incêndios, encontraram que a distância de áreas residenciais ou habitadas está entre as variáveis mais importantes para a predição desses eventos. Ainda em relação às variáveis que indicam a proximidade de atividades humanas, a distância de estradas, apesar de menor importância e contribuição no modelo, desempenha um papel fundamental nessa análise conjunta entre os fatores antrópicos, apontados como os principais fatores causadores de incêndios florestais (Nami *et al.*, 2018; Pang *et al.*, 2022; Celis *et al.*, 2023).

Relativo a pouca influência no processo de regressão da distância a estradas pode ser entendida pela natureza incompatível entre as duas variáveis (distância a estradas e densidade *kernel*). A distância a estrada perde o seu significado quando correlacionado com a saída de densidade kernel, em vez da localização do ponto de fogo (Amatulli *et al.*, 2006). A variável pode ser melhor analisada, por exemplo, considerando a distância absoluta dos pontos de fogo ou em termos de densidade da rede de estradas (Chuvieco *et al.*, 1999). Estas abordagens alternativas podem destacar a importância da variável como fator de risco de incêndio, com seu papel nas atividades operacionais de pré-supressão de incêndio.

A maior proximidade de áreas habitadas e de estradas apresentada em áreas mais susceptíveis a incêndios evidencia os impactos da urbanização nessas áreas. A maior concentração de atividades humanas nessas áreas ocasiona impactos socioambientais, refletidos não somente no aumento da ocorrência de incêndios, mas também em deslizamentos, enchentes, erosão, contaminação do lençol freático e desmatamento (Pasqualotto e Sena, 2017; Gama *et al.*, 2019). Em estudos sobre a dinâmica de focos de calor na APA Triunfo do Xingu, Rosan *et al.* (2017) e Dos Santos *et al.* (2023) corroboram que as maiores concentrações de focos de calor na APA ocorrem nas áreas de maior concentração populacional, nas frentes de ocupação e expansão econômica. Sousa *et al.* (2016), destacam ainda que a APA possui diversas áreas abertas ao longo de estradas, com áreas de floresta e terras agricultáveis em disputa entre agricultores familiares e fazendeiros, o que aumenta a pressão de

desmatamento sobre essas áreas, que possui relação direta com o aumento de queimadas, como apontado por Dos Santos *et al.* (2023).

O tipo de uso e cobertura da terra é outro fator com influência e contribuição consideráveis neste caso. Por refletir a interação de atividades humanas e meios naturais, sua influência e contribuição pode ter sido mascarada em outros elementos, mas ainda expressa informações importantes para o modelo de predição. A importância do uso da terra é apontada por Pourtaghi *et al.* (2016), Pourghasemi *et al.* (2020b) e Pham *et al.* (2020) como uma das variáveis mais importantes no processo de predição e mapeamento da susceptibilidade de incêndios.

## **6 CONSIDERAÇÕES FINAIS**

Este estudo demonstrou a capacidade de dois algoritmos ensemble, RF e XGB, de realizar com precisão, a predição de áreas susceptíveis a ocorrência de incêndios florestais na APA Triunfo do Xingu. O desempenho similar dos modelos indicou que ambos podem ser usados no mapeamento dessas áreas. O mapa baseado nos resultados do modelo XGB apresentou áreas de maior susceptibilidade nas porções centrais leste e oeste da unidade de conservação. As variáveis ambientais e socioeconômicas tiveram maior importância e contribuição no modelo, mostrando que áreas mais susceptíveis estão mais próximas de atividades humanas, recebem um menor volume anual de precipitação e possuem predominância de pastagens.

A comparação com outros estudos também permitiu demonstrar, nesses casos, um avanço dos modelos RF e XGB em relação ao CART, em termos de aumento da capacidade preditiva, mesmo que para isso seja necessário abrir mão de um conhecimento mais aprofundado sobre as estruturas de decisão internas do modelo. E ainda em relação aos dois modelos testados, mesmo que neste caso, não tenha ocorrido diferenças significativas na capacidade preditiva, ainda assim o modelo XGB apresentou um desempenho computacional superior ao modelo RF, sendo preferencial o seu uso por fatores como tempo de execução e escalabilidade.

Para o melhor do nosso conhecimento, este é o primeiro estudo sobre a aplicação de técnicas de ML na APA Triunfo do Xingu, contribuindo para a melhoria do entendimento da dinâmica de ocorrência de focos de queima nessa área protegida. Ainda assim, a utilização de dados com resolução espacial e temporal melhores, pode

ajudar ainda mais no entendimento e combate de incêndios florestais em áreas de conservação, evitando a perda dos recursos florestais e também de suas funções ecológicas.

Em relação a trabalhos futuros, é importante a continuidade em volume e frequência de estudos na região amazônica, englobando principalmente áreas de unidades de conservação. Outro foco a ser mais desenvolvido se refere à sensibilidade dos algoritmos com relação às variáveis, o que pode trazer novas contribuições para o entendimento sobre a interação de fatores físicos, ambientais e sociais sobre a ocorrência de incêndios.

Com novas tecnologias surgindo todos os dias, novos modelos e técnicas têm ganhado papel de destaque, principalmente quando nos referimos a modelos de DL. Novas pesquisas podem explorar melhor a capacidade desses modelos no contexto da predição de incêndios florestais. Novas variáveis com resoluções melhores também podem ter impacto significativo nesse processo, uma vez que com uma maior disponibilidade de dados com qualidade cada vez melhor pode ser capaz de melhorar a capacidade preditiva dos modelos e também gerar mapas com maior qualidade de detalhes.

Dessa forma, é de grande importância a realização de estudos que explorem novas abordagens, integrando diferentes fontes de dados e que conciliem as novas técnicas de inteligência artificial, ML e DL à expertise humana, para enfrentar esses desafios, protegendo tanto as florestas quanto as comunidades que dependem delas. Assim, podemos alavancar a importância da inteligência artificial, pavimentando o caminho para um futuro mais seguro e sustentável, onde a tecnologia se torna uma aliada vital na preservação dos ecossistemas naturais.

## REFERÊNCIAS

ABDOLLAHI, Masoud; ISLAM, Tanvir; GUPTA, Anil; HASSAN, Quazi K. An advanced forest fire danger forecasting system: Integration of remote sensing and historical sources of ignition data. **Remote Sensing**, v. 10, n. 6, p. 923, 2018. <https://doi.org/10.3390/rs10060923>

ABEDI GHESHLAGHI, Hassan; FEIZIZADEH, Bakhtiar; BLASCHKE, Thomas. GIS-based forest fire risk mapping using the analytical network process and fuzzy

logic. **Journal of Environmental Planning and Management**, v. 63, n. 3, p. 481-499, 2020. <https://doi.org/10.1080/09640568.2019.1594726>

ABID, Faroudja. A survey of machine learning algorithms based forest fires prediction and detection systems. **Fire technology**, v. 57, n. 2, p. 559-590, 2021. <https://doi.org/10.1007/s10694-020-01056-z>

ADITAMA, Tjandra Yoga. Impact of haze from forest fire to respiratory health: Indonesian experience. **Respirology**, v. 5, n. 2, p. 169-174, 2000. <https://doi.org/10.1046/j.1440-1843.2000.00246.x>

AGRAWAL, Nikita; NELSON, Peder V.; LOW, Russanne D. A Novel Approach for Predicting Large Wildfires Using Machine Learning towards Environmental Justice via Environmental Remote Sensing and Atmospheric Reanalysis Data across the United States. **Remote Sensing**, v. 15, n. 23, p. 5501, 2023. <https://doi.org/10.3390/rs15235501>

ALBAR, Israr; JAYA, I. Nengah Surati; SAHARJO, Bambang Hero; KUNCAHYO, Budi; VADREVU, Krishna Prasad. Spatio-temporal analysis of land and forest fires in Indonesia using MODIS active fire dataset. **Land-atmospheric research applications in South and Southeast Asia**, p. 105-127, 2018. [https://doi.org/10.1007/978-3-319-67474-2\\_6](https://doi.org/10.1007/978-3-319-67474-2_6)

ALDHYANI, Theyazn HH; AL-YAARI, Mohammed; ALKAHTANI, Hasan; MAASHI, Mashaël. Water quality prediction using artificial intelligence algorithms. **Applied Bionics and Biomechanics**, v. 2020, 2020. <https://doi.org/10.1155/2020/6659314>

ALENCAR, Ane. A.; BRANDO, Paulo. M.; ASNER, Gregory. P; PUTZ, Francis. E. Landscape fragmentation, severe drought, and the new Amazon Forest fire regime. **Ecological Applications**, 25: 1493-1505, 2015. <https://doi.org/10.1890/14-1528.1>

ALENCAR, Ane; SILVESTRINI, Rafaella; GOMES, Jarlene; SAVIAN, Gabriela. Amazônia em chamas: o novo e alarmante patamar do desmatamento na Amazônia. **Instituto de Pesquisa Ambiental da Amazônia-IPAM**, 2022. Disponível em: <https://static.poder360.com.br/2022/02/nota-tecnica-ipam-desmatamento-2-fev-2022.pdf>

ALMEIDA, Tarsis Esaú Gomes; FLORES, Maria do Socorro Almeida; SOBRINHO, Mário Vasconcellos. Mapeamento de Risco de Desastre por Incêndio Florestal na Amazônia: uma abordagem multifatorial no município de Moju (PA). **InterEspaço: Revista de Geografia e Interdisciplinaridade**, v. 6, n. 19, p. 2020. <https://doi.org/10.18764/2446-6549.202009>

ALVES, Raimundo Nonato Brabo; MODESTO JUNIOR, Moises de Souza. **Roça sem fogo: da tradição das queimadas para a agricultura sustentável na Amazônia**. Brasília: Embrapa. 184 p. 2020.

AMATULLI, Giuseppe; RODRIGUES, Maria João; TROMBETTI, Marco; LOVREGLIO, Raffaella. Assessing long-term fire risk at local scale by means of decision tree technique. **J Geophys Res** 111(4): 1-15. 2006. <https://doi.org/10.1029/2005JG000133>

ANDRADE, Caio F. de; DUARTE, João; BARBOSA, Maria Lucia Ferreira; ANDRADE, Melina de; OLIVEIRA, Raquel O. de; DELGADO, Rafael; PEREIRA, Marcos; BATISTA, Tays; TEODORO, Paulo. Fire outbreaks in extreme climate years in the State of Rio de Janeiro, Brazil. **Land Degradation & Development**, v. 30, n. 11, p. 1379-1389, 2019. <https://doi.org/10.1002/ldr.3327>

ANDRADE, Shirlei Cardoso; FERREIRA, Alex Franco. Mapeamento Geoecológico Da Susceptibilidade à Ocorrência De Incêndios No Parque Estadual Da Serra Da Concórdia–Valença RJ. **Revista Eletrônica TECCEN**, v. 12, n. 2, p. 45-58, 2019. <https://doi.org/10.21727/teccen.v12i2.1999>

ARAGÃO, Luiz; ANDERSON, Liana; FONSECA, Marisa; ROSAN, Thais; VEDOVATO, Laura; WAGNER, Fabien; SILVA, Camila; SILVA JUNIOR, Celso; ARAI, Egidio; AGUIAR, Ana; BARLOW, Jos; BERENGUER, Erika; DEETER, Merritt; DOMINGUES, Lucas; GATTI, Luciana; GLOOR, Manuel; MALHI, Yadvinder; MARENGO, Jose; MILLER, John; PHILLIPS, Oliver; SAATCHI, Sassan. 21st Century drought-related fires counteract the decline of Amazon deforestation carbon emissions. **Nature communications**, v. 9, n. 1, p. 536, 2018. <https://doi.org/10.1038/s41467-017-02771-y>

ARAÚJO, Elis; BARRETO, Paulo; BAIMA, Sara; GOMES, Mayara. **Unidades de conservação mais desmatadas da Amazônia Legal (2012–2015)**. Belém: Imazon, 2017.

ARLOT, Sylvain; CELISSE, Alain. A survey of cross-validation procedures for model selection. **Statistics Surveys**, v.4, p.40 - 79, 2010. <https://doi.org/10.1214/09-SS054>

AWAN, Abid Ali. **What is Bagging in Machine Learning?** A Guide with Examples. DataCamp, 2023. Disponível em: <https://www.datacamp.com/tutorial/what-bagging-in-machine-learning-a-guide-with-examples>. Acesso em: 14 maio 2024.

BACANI, Vitor Matheus. Geoprocessing applied to risk assessment of forest fires in the municipality of Bodoquena, Mato Grosso do Sul. **Revista Árvore**, v. 40, p. 1003-1011, 2016. <https://doi.org/10.1590/0100-67622016000600005>

BATISTA, Gustavo E. A. P. A.; MONARD, Maria Carolina. An analysis of four missing data treatment methods for supervised learning. **Applied artificial intelligence**, v. 17, n. 5-6, p. 519-533, 2003. <https://doi.org/10.1080/713827181>

BERLINCK, Christian N.; BATISTA, Eugênia K. L. Good fire, bad fire: It depends on who burns. **Flora**, v. 268, p. 151610, 2020. <https://doi.org/10.1016/j.flora.2020.151610>

BEVEN, Keith J.; KIRKBY, Michael J. A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du 83asin versant. **Hydrological sciences journal**, v. 24, n. 1, p. 43-69, 1979. <https://doi.org/10.1080/02626667909491834>

BILUCAN, Furkan; TEKE, Alihan; KAVZOGLU, Taskin. Susceptibility Mapping of Wildfires Using XGBoost, Random Forest and AdaBoost: A Case Study of

Mediterranean Ecosystem. In: **international conference on Mediterranean Geosciences Union**. Cham: Springer Nature Switzerland, p. 99-101. 2022. [http://dx.doi.org/10.1007/978-3-031-48715-6\\_22](http://dx.doi.org/10.1007/978-3-031-48715-6_22)

BRANDO, Paulo M.; PAOLUCCI, Lucas; UMMENHOFER, Caroline C.; ORDWAY, Elsa M.; HARTMANN, Henrik; CATTAU, Megan E.; RATTIS, Ludmila; MEDJIBE, Vincent; COE, Michael T.; BALCH, Jennifer. Droughts, wildfires, and forest carbon cycling: A pantropical synthesis. **Annual Review of Earth and Planetary Sciences**, v. 47, p. 555-581, 2019. <https://doi.org/10.1146/annurev-earth-082517-010235>

BREIMAN, L.; FRIEDMAN, J. H.; OLSHEN, R. A.; STONE, C. J. Classification and regression trees. Brooks. **Wadsworth and Brooks, Monterey, CA**, 1984. <https://doi.org/10.1201/9781315139470>

BREIMAN, Leo. Bagging predictors. **Machine learning**, v. 24, p. 123-140, 1996. <https://doi.org/10.1007/BF00058655>

BREIMAN, Leo. Random forests. **Machine learning**, v. 45, p. 5-32, 2001. <https://doi.org/10.1023/A:1010933404324>

BUI, Dieu Tien; BUI, Quang-Thanh; NGUYEN, Quoc-Phi; PRADHAN, Biswajeet; NAMPAK, Haleh; TRINH, Phan Trong. A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area. **Agricultural and forest meteorology**, v. 233, p. 32-44, 2017. <https://doi.org/10.1016/j.agrformet.2016.11.002>

CARRERO, Gabriel; SOARES, Pedro; MARINHO, Thiago. **Entendendo as queimadas e incêndios florestais em Apuí, Sul do Amazonas**. Manaus: Idesam, 2019.

CARVALHO, Aldani Braz; MOREIRA, Rodolfo Pragana; HERRERA, José Antônio. Aspectos da Dinâmica Climática de Altamira-PA. **REVISTA PERCURSO**, v. 14, n. 2, p. 23-34, 2022.

CASALICCHIO, Giuseppe; MOLNAR, Christoph; BISCHL, Bernd. Visualizing the feature importance for black box models. In: Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2018, Dublin, Ireland, September 10–14, 2018, Proceedings, Part I 18. **Springer International Publishing**, 2019. p. 655-670. [https://doi.org/10.1007/978-3-030-10925-7\\_40](https://doi.org/10.1007/978-3-030-10925-7_40)

CASTRO, Hermano Albuquerque de; GONÇALVES, Karen dos Santos; HACON, Sandra de Souza. Trend of mortality from respiratory disease in elderly and the forest fires in the state of Rondônia/Brazil period between 1998 and 2005. **Ciênc Saúde Coletiva**, v. 14, n. 6, p. 2083-2090, 2009. <https://doi.org/10.1590/S1413-81232009000600015>

CAÚLA, R. H., OLIVEIRA-JÚNIOR, J. F., LYRA, G. B., DELGADO, R. C., & HEILBRON FILHO, P. F. L. Overview of fire foci causes and locations in Brazil based on meteorological satellite data from 1998 to 2011. **Environmental Earth Sciences**, v. 74, p. 1497-1508, 2015. <https://doi.org/10.1007/s12665-015-4142-z>

CELIS, Nathalia; CASALLAS, Alejandro; LOPEZ-BARRERA, Ellie Anne; FELICIAN, Martina; DE MARCHI, Massimo; PAPPALARDO, Salvatore. Climate change, forest fires, and territorial dynamics in the Amazon rainforest: an integrated analysis for mitigation strategies. **ISPRS International Journal of Geo-Information**, v. 12, n. 10, p. 436, 2023. <https://doi.org/10.3390/ijgi12100436>

CHAN, Jireh Yi-Le et al. Mitigating the multicollinearity problem and its machine learning approach: a review. **Mathematics**, v. 10, n. 8, p. 1283, 2022. <https://doi.org/10.3390/math10081283>

CHEN, Tianqi; GUESTIN, Carlos. Xgboost: A scalable tree boosting system. In: **Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining**. 2016. p. 785-794. <https://doi.org/10.1145/2939672.2939785>

CHUVIECO E. Remote Sensing of Large Wildfires in the European Mediterranean Basin. Berlin: **Springer**, 212 p, 1999. <https://doi.org/10.1007/978-3-642-60164-4>

COSTAFREDA-AUMEDES, Sergi; COMAS, Carles; VEJA-GARCIA, Cristina. Human-caused fire occurrence modelling in perspective: A review. **International Journal of Wildland Fire**, v. 26, n. 12, p. 983-998, 2017. <https://doi.org/10.1071/WF17026>

COZMAN, Fabio Gagliardi; KAUFMAN, Dora. Viés no aprendizado de máquina em sistemas de inteligência artificial: a diversidade de origens e os caminhos de mitigação. **Revista USP**, n. 135, p. 195-210, 2022. <https://doi.org/10.11606/issn.2316-9036.i135p195-210>

DA SILVA, Davi Guimarães; GELLER, Marla Teresinha Barbosa; MOURA, Mauro Sérgio dos Santos; MENESES, Anderson Alvarenga de Moura. Performance evaluation of LSTM neural networks for consumption prediction. **E-Prime-Advances in Electrical Engineering, Electronics and Energy**, v. 2, p. 100030, 2022. <https://doi.org/10.1016/j.prime.2022.100030>

DANGETI, Pratap. **Statistics for Machine Learning: Techniques for exploring supervised, unsupervised, and reinforcement learning models with Python and R**. Packt Publishing Ltd, 2017.

DE BEM, Pablo Pozzobon; DE CARVALHO JUNIOR, Osmar Abílio; MATRICARDI, Eraldo Aparecido Trondoli; GUIMARÃES, Renato Fontes; GOMES, Roberto Arnaldo Trancoso. Predicting wildfire vulnerability using logistic regression and artificial neural networks: a case study in Brazil's Federal District. **International journal of wildland fire**, v. 28, n. 1, p. 35-45, 2018. <https://doi.org/10.1071/WF18018>

DE JESUS, Janisson B.; SANTANA, Ighor DM. Estimation of land surface temperature in caatinga area using Landsat 8 data. **Journal of Hyperspectral Remote Sensing**, v. 7, n. 3, p. 150-157, 2017. <https://doi.org/10.29150/jhrs.v7.3.p150-157>

DNIT – Departamento Nacional de Infraestrutura de Transportes. **Visualizador de Dados do DNITGeo**, 2022. Disponível em: <https://servicos.dnit.gov.br/vgeo/>

DOS SANTOS, Gabriel Garreto; NERIS, João Paulo Ferreira; COELHO, Roberta de Fátima Rodrigues. Dinâmica dos Focos de Calor na Área de Proteção Ambiental Triunfo do Xingu, Amazônia Paraense. **Revista GeoAmazônia**, v. 11, n. 22, p. 23-45, 2023. <http://dx.doi.org/10.18542/geo.v11i22.13770>

DOS SANTOS, Gabriel Garreto; NERIS, João Paulo Ferreira; DOS SANTOS, Maciel Garreto, DE FREITAS, Tatiana Pará Monteiro. Uso de geotecnologias na análise espacial dos focos de calor no município de São Félix do Xingu, Pará. **Geografia: Publicações Avulsas**, v. 2, n. 1, p. 395-419, 2020.

DUFF, Thomas J.; CAWSON, Jane G.; PENMAN, Trent D. Determining burnability: Predicting completion rates and coverage of prescribed burns for fuel management. **Forest ecology and management**, v. 433, p. 431-440, 2019. <https://doi.org/10.1016/j.foreco.2018.11.009>

EDGELEY, Catrin M.; PAVEGLIO, Travis B. Community recovery and assistance following large wildfires: The case of the Carlton Complex Fire. **International journal of disaster risk reduction**, v. 25, p. 137-146, 2017. <https://doi.org/10.1016/j.ijdrr.2017.09.009>

ELLWANGER, Joel Henrique; KULMANN-LEAL, Bruna; KAMINSKI, Valéria; VALVERDE-VILLEGAS; Jacqueline María; VEIGA, Ana Beatriz G. da; SPILKI, Fernando R.; FEARNSSIDE, Philip M.; CAESAR, Lílian; GIATTI, Leandro Luiz; WALLAU, Gabriel L.; ALMEIDA, Sabrina E.M.; BORBA, Mauro R.; HORA, Vanusa P. da; CHIES, José Artur B. Beyond diversity loss and climate change: Impacts of Amazon deforestation on infectious diseases and public health. **Anais da Academia Brasileira de Ciências**, v. 92, 2020. <https://doi.org/10.1590/0001-3765202020191375>

FACCHINELLI, Francesco; PAPPALARDO, Salvatore Eugenio; CODATO, Daniele; DIANTINI, Alberto; DELLA FERA, Giuseppe; CRESCINI, Edoardo; DE MARCHI, Massimo. Unburnable and unleakable carbon in western amazon: using VIIRS Nightfire data to map gas flaring and policy compliance in the Yasuní Biosphere Reserve. **Sustainability**, v. 12, n. 1, p. 58, 2019. <https://doi.org/10.3390/su12010058>

FERNANDES, Ana Sabrina Soares; TERCEIRO, Iago Braga. Queimadas e doenças respiratórias em crianças na cidade de Marabá, Pará/Fires and respiratory diseases in children in the city of Marabá, Pará Incendios y enfermedades respiratorias en niños de la ciudad de Marabá, Pará. **Brazilian Journal of Health Review**, v. 5, n. 1, p. 154-168, 2022. <https://doi.org/10.34119/bjhrv5n1-014>

FERREIRA, Joice; LENNOX, Gareth; GARDNER, Toby; THOMSON, James; BERENQUER, Erika; LEES, Alexander; MAC NALLY, Ralph; ARAGÃO, Luiz; FERRAZ, Silvio; LOUZADA, Julio; MOURA, Nárgila; OLIVEIRA, Victor; PARDINI, Renata; SOLAR, Ricardo; VIEIRA, Ima; BARLOW, Jos. Carbon-focused conservation may fail to protect the most biodiverse tropical forests. **Nature Climate Change**, v. 8, n. 8, p. 744-749, 2018. <https://doi.org/10.1038/s41558-018-0225-7>

FONSECA-MORELLO, Thiago; RAMOS, Rossano; STEIL, Lara; PARRY, Luke; BARLOW, Jos; MARKUSSON, Nils; FERREIRA, Amanda. Queimadas E Incêndios

Florestais Na Amazônia Brasileira: Porque As Políticas Públicas Têm Efeito Limitado? **Ambiente & sociedade**, v. 20, p. 19-38, 2017. <https://doi.org/10.1590/1809-4422asoc0232r1v2042017>

FUCHS, Vanessa Boanada. Expansão de fronteira impulsionada pela China na Amazônia: quatro eixos de pressão causados pela crescente demanda pelo comércio de soja. **Civitas-Revista de Ciências Sociais**, v. 20, p. 16-31, 2020. <https://doi.org/10.15448/1984-7289.2020.1.34656>

GABARDO, Gislaine; SARZEDAS, Carolina Galvão; DA SILVA, Henrique Luis. Queimadas na Amazônia brasileira: Brasil em chamas. 2021. **A educação ambiental em uma perspectiva interdisciplinar**. <https://doi.org/10.37885/200800872>

GAMA, Luana Helena; OLIVEIRA, Maria José; SILVA, Thaciane Christine; NEVES, Sandrya; DIAS, Gustavo. Dinâmica de Uso do Solo e sua Relação com os Focos de Calor na Área de Preservação Ambiental Triunfo do Xingu-PA. **Enciclopédia Biosfera**, v. 16, n. 29, 2019.

GOMES, Dennis dos Santos. Inteligência Artificial: conceitos e aplicações. **Revista Olhar Científico**, v. 1, n. 2, p. 234-246, 2010.

GOMES, Vitor; VIEIRA, Ima; SALOMÃO, Rafael; TER STEEGE, Hans. Amazonian tree species threatened by deforestation and climate change. **Nature Climate Change**, v. 9, n. 7, p. 547-553, 2019. <https://doi.org/10.1038/s41558-019-0500-2>

GONZÁLEZ, Sergio; GARCÍA, Salvador; DEL SER, Javier; ROKACH, Lior; HERRERA, Francisco. A practical tutorial on bagging and boosting based ensembles for machine learning: Algorithms, software tools, performance study, practical perspectives and opportunities. **Information Fusion**, v. 64, p. 205-237, 2020. <https://doi.org/10.1016/j.inffus.2020.07.007>

GRÜBLER, Murillo. **Entendendo o funcionamento de uma Rede Neural Artificial**, 2018. Disponível em: <https://medium.com/brasil-ai/entendendo-o-funcionamento-de-uma-rede-neural-artificial-4463fcf44dd0>

GUO, Rui; ZHAO, Zhiqian; WANG, Tao; LIU, Guangheng; ZHAO, Jingyi; GAO, Dianrong. Degradation state recognition of piston pump based on ICEEMDAN and XGBoost. **Applied Sciences**, v. 10, n. 18, p. 6593, 2020. <https://doi.org/10.3390/app10186593>

HANDORF, Elizabeth; YIN, Yinuo; SLIFKER, Michael; LYNCH, Shannon. Variable selection in social-environmental data: sparse regression and tree ensemble machine learning approaches. **BMC medical research methodology**, v. 20, p. 1-10, 2020. <https://doi.org/10.1186/s12874-020-01183-9>

HANG, Hoang Thi; MALLICK, Javed; ALGADHI, Saeed; ALI BINDAJAM, Ahmed; ABDO, Hazem Ghassan. Exploring Forest Fire Susceptibility and Management Strategies in Western Himalaya: Integrating Ensemble Machine Learning and eXplainable AI for Accurate Prediction and Comprehensive Analysis. **Environmental Technology & Innovation**, p. 103655, 2024. <https://doi.org/10.1016/j.eti.2024.103655>

HARRISON, Matt. **Machine Learning**: Guia de referência rápido. 1ª edição. São Paulo — SP: Novatec Editora Ltda, 2020.

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. **The elements of statistical learning: data mining, inference, and prediction**. New York: springer, 2009. <https://doi.org/10.1007/978-0-387-21606-5>

HAYKIN, Simon. **Neural networks: a comprehensive foundation**. Prentice Hall PTR, 2007. <https://dl.acm.org/doi/book/10.5555/1213811>

HO, Tin Kam. Random decision forests. In: **Proceedings of 3rd international conference on document analysis and recognition**. IEEE, 1995. p. 278-282. <https://doi.org/10.1109/ICDAR.1995.598994>

IANNACE, Gino; CIABURRO, Giuseppe; TREMATERRA, Amelia. Wind turbine noise prediction using random forest regression. **Machines**, v. 7, n. 4, p. 69, 2019. <https://doi.org/10.3390/machines7040069>

IBGE – Instituto Brasileiro de Geografia e Estatística. **Amazônia Legal**. Edição 2020. Acesso em 06 jul. 2023. Disponível em: <https://www.ibge.gov.br/geociencias/cartas-emapas/mapas-regionais/15819-amazonia-legal.html?=&t=o-que-e>

IBGE – Instituto Brasileiro de Geografia e Estatística. **Estrutura Territorial: Localidades**, 2010. Disponível em: <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/estrutura-territorial/27385-localidades.html>

INPE – Instituto Nacional de Pesquisas Espaciais, 2023. **Sistema de Monitoramento – Banco de Dados de Queimadas**. Disponível: <http://queimadas.dgi.inpe.br/queimadas/bdqueimadas>. Acesso: 06 jul. 2023.

JANG, Eunna; KANG, Yoojin; IM, Jungho; LEE, Dong-Woo; YOON, Jongmin. Detection and monitoring of forest fires using Himawari-8 geostationary satellite data in South Korea. **Remote Sensing**, v. 11, n. 3, p. 271, 2019. <https://doi.org/10.3390/rs11030271>

JENKS, George F.; CASPALL, Fred C. Error on choroplethic maps: definition, measurement, reduction. **Annals of the Association of American Geographers**, v. 61, n. 2, p. 217-244, 1971. <https://doi.org/10.1111/j.1467-8306.1971.tb00779.x>

JOHNSTON, Fay; PURDIE, Stuart; JALALUDIN, Bin; MARTIN, Kara; HENDERSON, Sarah; MORGAN, Geoffrey. Air pollution events from forest fires and emergency department attendances in Sydney, Australia 1996–2007: a case-crossover analysis. **Environmental health**, v. 13, p. 1-9, 2014. <https://doi.org/10.1186/1476-069X-13-105>

JUVANHOL, Ronie Silva; FIEDLER, Nilton Cesar; DOS SANTOS, Alexandre; PELUZIO, Telma; DA SILVA, Wellington; PINHEIRO, Christiano Jorge; DE SOUSA, Helbecy Cristino. 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, 2023. <https://doi.org/10.1590/0001-3765202320201039>

KALANTAR, Bahareh; UEDA, Naonori; IDREES, Mohammed; JANIZADEH, Saeid; AHMADI, Kourosh; SHABANI, Farzin. Forest fire susceptibility prediction based on machine learning models with resampling algorithms on remote sensing data. **Remote Sensing**, v. 12, n. 22, p. 3682, 2020. <https://doi.org/10.3390/rs12223682>

KHAN, Abdullah Ayub; LAGHARI, Asif Ali; AWAN, Shafique Ahmed. Machine learning in computer vision: a review. **EAI Endorsed Transactions on Scalable Information Systems**, v. 8, n. 32, p. e4-e4, 2021a. <https://doi.org/10.4108/eai.21-4-2021.169418>

KHAN, Muhammad Yaseen; QAYOOM, Abdul; NIZAMI, Muhammad Suffian; SIDDIQUI, Muhammad Shoab; WASI, Shaukat; RAAZI, Muhammad Khaliq-ur-Rahman. Automated prediction of Good Dictionary EXamples (GDEX): a comprehensive experiment with distant supervision, machine learning, and word embedding-based deep learning techniques. **Complexity**, v. 2021, p. 1-18, 2021b. <https://doi.org/10.1155/2021/2553199>

LADISLAU, Felipe Fernandes; SILVA, Jaqueline da Consolação; MOREIRA, Ana Paula Minelli; NASCIMENTO, Gabriel Luca; RUCHKYS, Úrsula de Azevedo. Análise Multicritério aplicada ao mapeamento de risco de incêndio na APA Sul RMBH. **Caderno de Geografia**, v. 31, n. 66, p. 667, 2021. <https://doi.org/10.5752/P.2318-2962.2021v31n66p667>

LATORRE, Natalia Salazar; ARAGÃO, Luiz Eduardo Oliveira e Cruz; ANDERSON, Liana; ANDERE, Luaê; DUARTE, Valdete; ARAI, Egídio; LIMA, André. Impactos de queimadas sobre diferentes tipos de cobertura da terra no leste da Amazônia legal brasileira. **Revista Brasileira de Cartografia**, v. 69, n. 1, p. 179-192, 2017. <https://doi.org/10.14393/rbcv69n1-44039>

LI, Yudong; FENG, Zhongke; CHEN, Shilin; ZHAO, Ziyu; WANG, Fengge. Application of the artificial neural network and support vector machines in forest fire prediction in the Guangxi autonomous region, China. **Discrete Dynamics in Nature and Society**, v. 2020, p. 1-14, 2020. <https://doi.org/10.1155/2020/5612650>

LIBONATI, Renata; DACAMARA, Carlos; SETZER, Alberto; MORELLI, Fabiano; MELCHIORI, Arturo. An algorithm for burned area detection in the Brazilian Cerrado using 4  $\mu$ m MODIS imagery. **Remote sensing**, v. 7, n. 11, p. 15782-15803, 2015. <https://doi.org/10.3390/rs71115782>

LOUDERMILK, E. Louise; O'BRIEN, Joseph; GOODRICK, Scott; LINN, Rodman; SKOWRONSKI, Nicholas; HIERS, J. Kevin. Vegetation's influence on fire behavior goes beyond just being fuel. **Fire Ecology**, v. 18, n. 1, p. 9, 2022. <https://doi.org/10.1186/s42408-022-00132-9>

LOZANO, F. Javier; SUÁREZ-SEOANE, S; KELLY, M; LUIS, E. 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**, v. 112, n. 3, p. 708-719, 2008. <https://doi.org/10.1016/j.rse.2007.06.006>

LUDERMIR, Teresa Bernarda. Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências. **Estudos Avançados**, v. 35, p. 85-94, 2021. <https://doi.org/10.1590/s0103-4014.2021.35101.007>

MANN, Henry B.; WHITNEY, Donald R. On a test of whether one of two random variables is stochastically larger than the other. **The annals of mathematical statistics**, p. 50-60, 1947. <http://dx.doi.org/10.1214/aoms/1177730491>

MAPBIOMAS PROJECT. Collection 8 of the Annual Land Cover and Land Use Maps of Brazil (1985-2022). MapBiomass Data, V1. 2023. <https://doi.org/10.58053/MapBiomass/VJJCCL>

MARCÍLIO, Wilson E.; ELER, Danilo M. From explanations to feature selection: assessing SHAP values as feature selection mechanism. In: 2020 33rd SIBGRAPI conference on Graphics, Patterns and Images (SIBGRAPI). **IEEE**, 2020. p. 340-347. <https://doi.org/10.1109/SIBGRAPI51738.2020.00053>

MARTORANO, Lucieta Guerreiro. **Estudos climáticos do estado do Pará, classificação climática (Koppen) e deficiência hídrica (Thorntwaite, Mather)**. SUDAM: EMBRAPA-SNLCS, 1993.

MCCARTHY, John et al. **What is artificial intelligence**. 2007. Disponível em: <<https://www-formal.stanford.edu/jmc/whatisai.pdf>>

MESSIAS, Cassiano Gustavo; FERREIRA, Marcos César. Modelo Geoespacial para a Identificação de Áreas com Perigo de Propagação de Queimadas no Parque Nacional da Serra da Canastra. **Revista do Departamento de Geografia**, v. 38, p. 154-168, 2019. <https://doi.org/10.11606/rdg.v38i1.153493>

MICHAEL, Yaron; HELMAN, David; GLICKMAN, Oren; GABAY, David; BRENNER, Steve; LENSKY, Itamar. 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, 2021. <https://doi.org/10.1016/j.scitotenv.2020.142844>

MMA – Ministério do Meio Ambiente. **Cadastro Nacional de Unidades de Conservação (CNUC)**. Departamento de Áreas Protegidas, 2022. Disponível em: <https://cnuc.mma.gov.br/powerbi>

MOHAJANE, Meriame; COSTACHE, Romulus; KARIMI, Firoozeh; PHAM, Quoc Bao; ESSAHLAOU, Ali; NGUYEN, Hoang; LANEVE, Giovanni; OUDIJA, Fatiha. Application of remote sensing and machine learning algorithms for forest fire mapping in a Mediterranean area. **Ecological Indicators**, v. 129, p. 107869, 2021. <https://doi.org/10.1016/j.ecolind.2021.107869>

MOREIRA, Pedro Augusto Gonzaga; MENDES, Thiago Augusto; SANTOS, Diego Fonseca dos. Avaliação de locais potenciais para instalação de torres de observação para prevenção de risco de incêndios florestais. **Ciência Florestal**, v. 30, p. 1266-1282, 2020. <https://doi.org/10.5902/1980509839686>

NAMI, M. H; JAAFARI, A; FALLAH, M; NABIUNI, S. Spatial prediction of wildfire probability in the Hyrcanian ecoregion using evidential belief function model and GIS. **International journal of environmental science and technology**, v. 15, p. 373-384, 2018. <https://doi.org/10.1007/s13762-017-1371-6>

NEARY, G. D.; LEONARD, M. J. Effects of fire on grassland soils and water: A review. **Grasses and grassland aspects**, p. 1-22, 2020. <http://dx.doi.org/10.5772/intechopen.90747>

NEPSTAD, Daniel; VERSSIMO, Adalberto; ALENCAR, Ane; NOBRE, Carlos; LIMA, Eirivelthon; LEFEBVRE; Paul; SCHLESINGER, Peter; POTTER, Christopher; MOUTINHO, Paulo; MENDOZA, Elsa; COCHRANE, Mark; BROOKS, Vanessa. Large-scale impoverishment of Amazonian forests by logging and fire. **Nature**, v. 398, n. 6727, p. 505-508, 1999. <https://doi.org/10.1038/19066>

NEUHAUSER, Markus. **Nonparametric statistical tests: A computational approach**. CRC Press, 248 p. 2011. <https://doi.org/10.1201/b11427>

NGARAMBE, Jack; YUN, Geun Young; SANTAMOURIS, Mat. The use of artificial intelligence (AI) methods in the prediction of thermal comfort in buildings: Energy implications of AI-based thermal comfort controls. **Energy and Buildings**, v. 211, p. 109807, 2020. <https://doi.org/10.1016/j.enbuild.2020.109807>

NÓBREGA, Lawrence Oliveira de; LAZZARINI, Gustavo Maximiano Junqueira; VIOLA, Marcelo Ribeiro; BATISTA, Antônio Carlos; CARVALHO, Edmar Vinícius de; GIONGO, Marcos. Forest fire susceptibility index for assessing the history of fire occurrences in the indigenous land of Kraholândia, Brazil. **Advances in Forestry Science**, v. 5, n. 2, p. 325-332, 2018. <https://doi.org/10.34062/afs.v5i2.5841>

NORVIG, Peter; RUSSELL, Stuart. Inteligência artificial. **Rio de Janeiro: Grupo GEN**, 2013.

OLIVEIRA, Víncler Fernandes Ribeiro; DA SILVA, Edson Rodrigo dos Santos; DA SILVA, Bruno Henrique Machado; VICK, Eirivelton Pereira; LIMA, César Gustavo da Rocha; BACANI Vitor Matheus. Geoprocessamento aplicado ao mapeamento de risco a incêndios. **Revista Brasileira de Geografia Física**, v. 13, n. 03, p. 1194-1212, 2020. <https://doi.org/10.26848/rbgf.v13.3.p1194-1212>

OZENEN KAVLAK, Mehtap; CABUK, Saye Nihan; CETIN, Mehmet. Development of forest fire risk map using geographical information systems and remote sensing capabilities: Ören case. **Environmental Science and Pollution Research**, v. 28, n. 25, p. 33265-33291, 2021. <https://doi.org/10.1007/s11356-021-13080-9>

PAHLOW, Markus; KLEISSL, Jan; PARLANGE, Marc B. Atmospheric boundary-layer structure observed during a haze event due to forest-fire smoke. **Boundary-layer meteorology**, v. 114, p. 53-70, 2005. <https://doi.org/10.1007/s10546-004-6350-z>

PANG, Yongqi; LI, Yudong; FENG, Zhongke; FENG, Zemin; ZHAO, Ziyu; CHEN, Shilin; ZHANG, Hanyue. Forest fire occurrence prediction in China based on machine

learning methods. **Remote Sensing**, v. 14, n. 21, p. 5546, 2022. <https://doi.org/10.3390/rs14215546>

PASQUALOTTO, N.; SENA, M. M. Impactos ambientais urbanos no Brasil e os caminhos para cidades sustentáveis. **Revista Educação Ambiental em Ação**, v. 61, 2017.

PHAM, Binh Thai; JAAFARI, Abolfazl; AVAND, Mohammadtaghi; AL-ANSARI, Nadhir; DU, Tran Dinh; HAI YEN, Hoang Phan; VAN PHONG, Tran; NGUYEN, Duy Huu; VAN LE, Hiep; MAFI-GHOLAMI, Davood; PRAKASH, Indra; THUY, Hoang Thi; TUYEN, Tran Thi. Performance evaluation of machine learning methods for forest fire modeling and prediction. **Symmetry**, v. 12, n. 6, p. 1022, 2020. <https://doi.org/10.3390/sym12061022>

PHELPS, Nathan; WOOLFORD, Douglas G. 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, 2021. <https://doi.org/10.1071/WF20139>

PORTER, Michael D.; REICH, Brian J. Evaluating temporally weighted kernel density methods for predicting the next event location in a series. **Annals of GIS**, v. 18, n. 3, p. 225-240, 2012. <https://doi.org/10.1080/19475683.2012.691904>

POURGHASEMI, Hamid Reza; KARIMINEJAD, Narges; AMIRI, Mahdis; EDALAT, Mohsen; ZARAFSHAR, Mehrdad; BLASCHKE, Thomas; CERDA, Artemio. Assessing and mapping multi-hazard risk susceptibility using a machine learning technique. **Scientific reports**, v. 10, n. 1, p. 3203, 2020a. <https://doi.org/10.1038/s41598-020-60191-3>

POURGHASEMI, Hamid Reza; GAYEN, Amiya; LASAPONARA, Rosa; TIEFENBACHER, John. Application of learning vector quantization and different machine learning techniques to assessing forest fire influence factors and spatial modelling. **Environmental research**, v. 184, p. 109321, 2020b. <https://doi.org/10.1016/j.envres.2020.109321>

POURTAGHI, Zohre Sadat; POURGHASEMI, Hamid Reza; ARETANO, Roberta; SEMERARO, Teodoro. Investigation of general indicators influencing on forest fire and its susceptibility modeling using different data mining techniques. **Ecological indicators**, v. 64, p. 72-84, 2016. <https://doi.org/10.1016/j.ecolind.2015.12.030>

POURTAGHI, Zohre Sadat; POURGHASEMI, Hamid Reza; ROSSI, Mauro. Forest fire susceptibility mapping in the Minudasht forests, Golestan province, Iran. **Environmental Earth Sciences**, v. 73, n. 4, p. 1515-1533, 2015. <https://doi.org/10.1007/s12665-014-3502-4>

RAHMAN, Shoumik; MESEV, Victor. Change vector analysis, tasseled cap, and NDVI-NDMI for measuring land use/cover changes caused by a sudden short-term severe drought: 2011 Texas event. **Remote sensing**, v. 11, n. 19, p. 2217, 2019. <https://doi.org/10.3390/rs11192217>

RASEKHSCHAFFE, Keywan Christian; JONES, Robert C. Machine learning for stock selection. **Financial Analysts Journal**, v. 75, n. 3, p. 70-88, 2019. <https://doi.org/10.1080/0015198X.2019.1596678>

REBOITA, M. S.; GAN, M. A.; ROCHA, R. P.; AMBRIZZI, T. Regimes de precipitação na América do Sul: uma revisão bibliográfica. **Revista Brasileira de Meteorologia**, v. 25, n. 2, p. 185-204, 2010. <https://doi.org/10.1590/s0102-77862010000200004>

REDDINGTON, C. L; BUTT, E. W; RIDLEY, D. A; ARTAXO, P; MORGAN, W. T; SPRACKLEN, D. V. Air quality and human health improvements from reductions in deforestation-related fire in Brazil. **Nature Geoscience**, v. 8, n. 10, p. 768-771, 2015. <https://doi.org/10.1038/ngeo2535>

RIBEIRO, Luciene; SOARES, Ronaldo Viana; BEPLLER, Michele. Mapeamento do risco de incêndios florestais no município de Novo Mundo, Mato Grosso, Brasil. **Cerne**, v. 18, p. 117-126, 2012. <https://doi.org/10.1590/S0104-77602012000100014>

RIZZATTI, Maurício; BATISTA, Natália Lampert; SPODE, Pedro Leonardo Cezar; ERTHAL, Douglas Bouvier; FARIA, Rivaldo Mauro; SCOTTI, Anderson Augusto Volpato; TRENTIN, Romario; PETSCH, Carina; COSTA, Iago Turba; QUOOS, João Henrique. Mapeamento da COVID-19 por meio da densidade de Kernel. **Metodologias e Aprendizado**, v. 3, p. 44-53, 2020. <https://doi.org/10.21166/metapre.v3i0.1312>

RODRIGUES, Jeferson Botelho; SOARES DA SILVA, Débora Danna; NEVES SALES, Layse Lorena; NEGRÃO DE FREITAS, Stephanie Jael; LEITE CARVALHO CABRAL, Ana Carla. Análise de focos de queimadas no município de Balsas/MA. **Nature and Conservation**, v. 13, n. 3, p. 146-151, 2020. <https://doi.org/10.6008/CBPC2318-2881.2020.003.0015>

RODRIGUES, Julia; LIBONATI, Renata; PEREIRA, Allan; NOGUEIRA, Joana; SANTOS, Filipe; PERES, Leonardo; SANTA ROSA, Ananda; SCHROEDER, Wilfrid; PEREIRA, José; GIGLIO, Louis; TRIGO, Isabel; SETZER, Alberto. How well do global burned area products represent fire patterns in the Brazilian Savannas biome? An accuracy assessment of the MCD64 collections. **International Journal of Applied Earth Observation and Geoinformation**, v. 78, p. 318-331, 2019. <https://doi.org/10.1016/j.jag.2019.02.010>

ROKACH, Lior. **Ensemble Learning: Pattern classification using ensemble methods**. World scientific, 2019. <https://doi.org/10.1142/11325>

ROSAN, Thais Michele; ANDERSON, Liana O.; VEDOVATO, Laura. Avaliação da origem de focos de calor em anos de extremos climáticos na Amazônia brasileira. **Revista Brasileira de Cartografia**, v. 69, n. 4, p. 731-741, 2017. <https://doi.org/10.14393/rbcv69n4-44331>

RUBÍ, Jesús N. S; DE CARVALHO, Paulo H. P; GONDIM, Paulo R. L. 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, 2023. <https://doi.org/10.1016/j.engappai.2022.105649>

SÁ, Tatiana Deane de Abreu; KATO, Osvaldo Ryohei; CARVALHO, Claudio José Reis de; FIGUEIREDO, Ricardo de Oliveira. Queimar ou não queimar?: De como produzir na Amazônia sem queimar . **Revista USP**, São Paulo, Brasil, n. 72, p. 90–97, 2007. <https://doi.org/10.11606/issn.2316-9036.v0i72p90-97>

SALES, Gil Mendes; PEREIRA, Jorge Luis Gavina; THALÊS, Marcelo Cordeiro; POCCARD-CHAPUIS, René; ALMEIDA, Arlete Silva. Emprego dos focos de calor na avaliação das áreas queimadas e em incêndios florestais em Paragominas, Pará, Brasil. **Boletim do Museu Paraense Emílio Goeldi**. Ciências Naturais 14(1): 55-77, 2019. <https://doi.org/10.46357/bcnaturais.v14i1.140>

SANTOS, H.G., JACOMINE, P.K.T., ANJOS, L.H.C., OLIVEIRA, V.A., LUMBRERAS, J.F., COELHO, M.R., ALMEIDA, J.A., ARAUJO FILHO, K.C., OLIVEIRA, A.J.B., CUNHA, T.J.F. **Sistema Brasileiro de Classificação de Solos**. Brasília, DF: Embrapa, 2018. Disponível em: <https://www.embrapa.br/solos/sibcs>

SANTOS, Thiago Oliveira; ANDRADE FILHO, Valdir Soares; ROCHA, Vinícius Machado; MENEZES, Janaína de Souza. Os impactos do desmatamento e queimadas de origem antrópica sobre o clima da Amazônia brasileira: um estudo de revisão. **Revista Geográfica Acadêmica**, v. 11, n. 2, p. 157-181, 2017.

SCHAPIRE, Robert E.; FREUND, Yoav. Boosting: Foundations and algorithms. **Kybernetes**, v. 42, n. 1, p. 164-166, 2013. <https://doi.org/10.1108/03684921311295547>

SCHROEDER, Wilfrid; ALENCAR, Ane; ARIMA, Eugenio; SETZER, Alberto. The spatial distribution and interannual variability of fire in Amazonia. **Washington DC American Geophysical Union Geophysical Monograph Series**, v. 186, p. 43-60, 2009. <https://doi.org/10.1029/2008GM000723>

SCIKIT-LEARN. **Cross-validation**: evaluating estimator performance. Disponível em: [https://scikit-learn.org/stable/modules/cross\\_validation.html](https://scikit-learn.org/stable/modules/cross_validation.html). Acesso em: 7 jun. 2024.

SEDDOUKI, M; BENAYAD, M; AAMIR, Z; TAHIRI, M; MAANAN, M; RHINANE, H. Using Machine Learning Coupled with Remote Sensing for Forest Fire Susceptibility Mapping. Case Study Tetouan Province, Northern Morocco. **The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, v. 48, p. 333-342, 2023. <http://dx.doi.org/10.5194/isprs-archives-XLVIII-4-W6-2022-333-2023>

SHMUEL, Assaf; HEIFETZ, Eyal. A machine-learning approach to predicting daily wildfire expansion rate. **Fire**, v. 6, n. 8, p. 319, 2023. <https://doi.org/10.3390/fire6080319>

SILVA JUNIOR, Celso; ANDERSON, Liana; SILVA, Alindomar; ALMEIDA, Catherine; DALAGNOL, Ricardo; PLETSCHE, Mikhaela; PENHA, Thales; PALOSCHI, Rennan; ARAGÃO, Luiz. Fire responses to the 2010 and 2015/2016 Amazonian

droughts. **Frontiers in Earth Science**, p. 97, 2019. <https://doi.org/10.3389/feart.2019.00097>

SIMON, Marcelo; GREYER, Rosaura; DE QUEIROZ, Luciano; SKEMA, Cynthia; PENNINGTON, R. Toby; HUGHES, Colin. Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by in situ evolution of adaptations to fire. **Proceedings of the National Academy of Sciences**, v. 106, n. 48, p. 20359-20364, 2009. <https://doi.org/10.1073/pnas.090341010>

SOUSA, Romier da Paixão; DA SILVA, Ruth Corrêa; MIRANDA, Katiuscia; NETO, Manuel Amaral. **Governança socioambiental na Amazônia: Agricultura familiar e os desafios para a sustentabilidade em São Félix do Xingu-Pará**. Instituto Internacional de Educação do Brasil, 2016.

SOUZA JR, Carlos; FONSECA, Antônio; NUNES, Sâmia; SALOMÃO, Rodney; RIBEIRO, Júlia; MARTINS, Heron. Desmatamento em Áreas Protegidas. **O estado das áreas protegidas, Imazon**, 2018.

SOUZA, Alana Almeida de; OVIEDO, Antonio; DOS SANTOS, Tiago Moreira. Impactos na qualidade do ar e saúde humana relacionados ao desmatamento e queimadas na Amazônia Legal brasileira. **Instituto Socioambiental: São Paulo, SP, Brazil**, p. 21, 2020.

SPEARMAN, Charles. The Proof and Measurement of Association between Two Things. **The American Journal of Psychology**. 15 (1): 72–101, 1904. <https://doi.org/10.2307/1412159>

SYNERGIA SOCIOAMBIENTAL. **Focos de Calor, Áreas Protegidas na Amazônia Legal**. Coordenação de Geoprocessamento, Gerência de Estudos e Pesquisas, 36 p. 2022. Disponível em: [https://www.synergiaconsultoria.com.br/wp-content/uploads/2022/03/Relatorio\\_FocosCalor\\_AmazoniaLegal\\_VFinal\\_fev2022\\_editorada.pdf](https://www.synergiaconsultoria.com.br/wp-content/uploads/2022/03/Relatorio_FocosCalor_AmazoniaLegal_VFinal_fev2022_editorada.pdf)

TALAVIYA, Tanha; SHAH, Dhara; PATEL, Nivedita; YAGNIK, Hiteshri; SHAH, Manan. Implementation of artificial intelligence in agriculture for optimization of irrigation and application of pesticides and herbicides. **Artificial Intelligence in Agriculture**, v. 4, p. 58-73, 2020. <https://doi.org/10.1016/j.aiia.2020.04.002>

THEODORIDIS, Georgios; TSADIRAS, Athanasios. Applying machine learning techniques to predict and explain subscriber churn of an online drug information platform. **Neural Computing and Applications**, v. 34, n. 22, p. 19501-19514, 2022. <https://doi.org/10.1007/s00521-022-07603-9>

TIAN, Yuping; WU, Zechuan; LI, Mingze; WANG, Bin; ZHANG, Xiaodi. Forest fire spread monitoring and vegetation dynamics detection based on multi-source remote sensing images. **Remote Sensing**, v. 14, n. 18, p. 4431, 2022. <https://doi.org/10.3390/rs14184431>

TONINI, Marj; D'ANDREA, Mirko; BIONDI, Guido; ESPOSTI, Silvia Degli; TRUCCHIA, Andrea; FIORUCCI, Paolo. 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, 2020. <https://doi.org/10.3390/geosciences10030105>

TORRES, Fillipe Tamiozzo Pereira; ROQUE, Mariane Paulina Batalha; LIMA, Gumercindo Souza; MARTINS, Sebastião Venâncio; DE FARIA, André Luiz Lopes. Mapeamento do risco de incêndios florestais utilizando técnicas de geoprocessamento. **Floresta e Ambiente**, v. 24, 2017. <https://doi.org/10.1590/2179-8087.025615>

TURING, Alan M. **Computing machinery and intelligence**. Springer Netherlands, 2009. [https://doi.org/10.1007/978-1-4020-6710-5\\_3](https://doi.org/10.1007/978-1-4020-6710-5_3)

VADREVU, Krishna Prasad; EATURU, Anuradha; BADARINATH, KVSc. Fire risk evaluation using multicriteria analysis—a case study. **Environmental monitoring and assessment**, v. 166, p. 223-239, 2010. <https://doi.org/10.1007/s10661-009-0997-3>

VASCONCELOS, Sumaia Saldanha; FEARNESIDE, PHILIP MARTIN; GRAÇA, Paulo Maurício Lima de Alencastro, DIAS, David Valentim; CORREIA, Francis Wagner Silva. Variability of vegetation fires with rain and deforestation in Brazil's state of Amazonas. **Remote Sensing of Environment**, v. 136, p. 199-209, 2013. <https://doi.org/10.1016/j.rse.2013.05.005>

VENKATESH, K.; PREETHI, K.; RAMESH, H. Evaluating the effects of forest fire on water balance using fire susceptibility maps. **Ecological Indicators**, v. 110, p. 105856, 2020. <https://doi.org/10.1016/j.ecolind.2019.105856>

VIGANÓ, Hevelyne Henn da Gama; DE SOUZA, Celso Correia; CRISTALDO, Marcia Ferreira; DE JESUS, Leandro. Redes neurais artificiais na previsão de queimadas e incêndios no Pantanal. **Revista Brasileira de Geografia Física**, v. 10, n. 05, p. 1355-1367, 2017. <https://doi.org/10.26848/rbgf.v.10.p1355-1367>

WILCOXON, Frank; KATTI, S. K; WILCOX, Roberta. Critical values and probability levels for the Wilcoxon rank sum test and the Wilcoxon signed rank test. **Selected tables in mathematical statistics**, v. 1, p. 171-259, 1970.

WITTEN, Ian H.; FRANK, Eibe. Data mining: practical machine learning tools and techniques with Java implementations. **Acm Sigmod Record**, v. 31, n. 1, p. 76-77, 2002. <https://doi.org/10.1145/507338.507355>

XIE, Ying; PENG, Minggang. Forest fire forecasting using ensemble learning approaches. **Neural Computing and Applications**, v. 31, n. 9, p. 4541-4550, 2019. <https://doi.org/10.1007/s00521-018-3515-0>

YAN, Hui; LIU, Guixiang. Fire's effects on grassland restoration and biodiversity conservation. **Sustainability**, v. 13, n. 21, p. 12016, 2021. <https://doi.org/10.3390/su132112016>

ZHANG, Peng; GUO, Zhiling, ULLAH, Sami; MELAGRAKI, Georgia; AFANTITIS, Antreas; LYNCH, Iseult. Nanotechnology and artificial intelligence to enable

sustainable and precision agriculture. **Nature Plants**, v. 7, n. 7, p. 864-876, 2021.  
<https://doi.org/10.1038/s41477-021-00946-6>

## APÊNDICE A – ARTIGO APRESENTADO NO XXVI ENCONTRO NACIONAL DE MODELAGEM COMPUTACIONAL

XXVI ENMC, XIV ECTM  
25 a 27 de outubro de 2023



### AVALIAÇÃO EXPLORATÓRIA DO ALGORITMO RANDOM FOREST PARA PREVISÃO DE INCÊNDIOS FLORESTAIS

**Kemuel Maciel Freitas**<sup>1,2</sup> – kemuel.macielf@gmail.com

**Carlos Henrique Pinto da Silveira**<sup>3</sup> – carlos.silveira@discente.ufopa.edu.br

**Christiano Jorge Gomes Pinheiro**<sup>4</sup> – christiano.pinheiro@ufes.br

**Jefferson Rossy Pereira da Silva**<sup>1,2</sup> – j.rossy95@gmail.com

**Ronie Silva Juvanhol**<sup>1,5</sup> – roniejuvanhol@ufpi.edu.br

**Anderson Alvarenga de Moura Meneses**<sup>1,2</sup> – anderson.meneses@ufopa.edu.br

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

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

<sup>3</sup> Laboratório de Modelagem Computacional, Instituto de Engenharia e Geociências, Universidade Federal do Oeste do Pará - Santarém, PA, Brasil

<sup>4</sup> Departamento de Engenharia Rural, Universidade Federal do Espírito Santo - Alegre, ES, Brasil

<sup>5</sup> 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

## APÊNDICE B – RESULTADOS DO GRIDSEARCHCV PARA OS ALGORITMOS RANDOM FOREST E XGBOOST

RANDOM FOREST									
<i>n_estimators</i> = 100 <i>max_depth</i> = None		<i>n_estimators</i> = 150 <i>max_depth</i> = None		<i>n_estimators</i> = 200 <i>max_depth</i> = None		<i>n_estimators</i> = 250 <i>max_depth</i> = None			
Mean Fit Time = 141.8s		Mean Fit Time = 208.6s		Mean Fit Time = 265.5s		Mean Fit Time = 343.7s			
RMSE		RMSE		RMSE		RMSE			
CROSS-VALIDATION (K = 10)	SPLIT 1	37.3368	SPLIT 1	36.9686	SPLIT 1	36.3823	SPLIT 1	36.3287	
	SPLIT 2	34.6631	SPLIT 2	34.3985	SPLIT 2	34.4831	SPLIT 2	34.9094	
	SPLIT 3	34.9594	SPLIT 3	35.0900	SPLIT 3	34.3429	SPLIT 3	36.1644	
	SPLIT 4	35.9942	SPLIT 4	36.2170	SPLIT 4	36.0847	SPLIT 4	36.9075	
	SPLIT 5	37.0002	SPLIT 5	37.2906	SPLIT 5	37.3249	SPLIT 5	36.8883	
	SPLIT 6	36.5355	SPLIT 6	36.7784	SPLIT 6	36.6378	SPLIT 6	38.0541	
	SPLIT 7	37.7228	SPLIT 7	37.9195	SPLIT 7	37.6419	SPLIT 7	36.2325	
	SPLIT 8	36.7531	SPLIT 8	36.6966	SPLIT 8	36.4337	SPLIT 8	37.2502	
	SPLIT 9	37.2299	SPLIT 9	37.0380	SPLIT 9	37.1917	SPLIT 9	34.8005	
	SPLIT 10	35.1722	SPLIT 10	34.5380	SPLIT 10	35.0259	SPLIT 10	36.2372	
MEAN	36.3367	MEAN	36.2935	MEAN	36.1549	MEAN	36.2372		
STD. DV.	1.0268	STD. DV.	1.1477	STD. DV.	1.1124	STD. DV.	1.0489		
MEDIAN	36.6443	MEDIAN	36.7375	MEDIAN	36.4080	MEDIAN	36.2829		
MIN.	34.6631	MIN.	34.3985	MIN.	34.3429	MIN.	34.8005		
MAX.	37.7228	MAX.	37.9195	MAX.	37.6419	MAX.	38.0541		

RANDOM FOREST								
<i>n_estimators</i> = 100 <i>max_depth</i> = 10		<i>n_estimators</i> = 150 <i>max_depth</i> = 10		<i>n_estimators</i> = 200 <i>max_depth</i> = 10		<i>n_estimators</i> = 250 <i>max_depth</i> = 10		
Mean Fit Time = 73.4s		Mean Fit Time = 109.9s		Mean Fit Time = 146.6s		Mean Fit Time = 187.0s		
RMSE		RMSE		RMSE		RMSE		
CROSS-VALIDATION (K = 10)	SPLIT 1	172.3661	CROSS-VALIDATION (K = 10)	SPLIT 1	173.1375	CROSS-VALIDATION (K = 10)	SPLIT 1	173.4452
	SPLIT 2	171.7553		SPLIT 2	172.2495		SPLIT 2	170.9267
	SPLIT 3	171.5507		SPLIT 3	171.0269		SPLIT 3	172.0383
	SPLIT 4	169.6531		SPLIT 4	170.5883		SPLIT 4	170.1940
	SPLIT 5	174.8525		SPLIT 5	172.8634		SPLIT 5	173.5682
	SPLIT 6	170.5321		SPLIT 6	170.5792		SPLIT 6	169.8410
	SPLIT 7	171.3632		SPLIT 7	171.1474		SPLIT 7	170.7861
	SPLIT 8	169.6044		SPLIT 8	168.7655		SPLIT 8	168.6380
	SPLIT 9	170.6938		SPLIT 9	171.6037		SPLIT 9	171.3760
	SPLIT 10	168.7768		SPLIT 10	168.8744		SPLIT 10	168.2366
MEAN	171.1148	MEAN	171.0836	MEAN	171.0223	MEAN	170.9050	
STD. DV.	1.6309	STD. DV.	1.4063	STD. DV.	1.8395	STD. DV.	1.7018	
MEDIAN	171.0285	MEDIAN	171.0871	MEDIAN	171.0442	MEDIAN	170.8564	
MIN.	168.7768	MIN.	168.7650	MIN.	168.0388	MIN.	168.2366	
MAX.	174.8525	MAX.	173.1375	MAX.	174.4477	MAX.	173.5682	

RANDOM FOREST									
<i>n_estimators</i> = 100 <i>max_depth</i> = 20		<i>n_estimators</i> = 150 <i>max_depth</i> = 20		<i>n_estimators</i> = 200 <i>max_depth</i> = 20		<i>n_estimators</i> = 250 <i>max_depth</i> = 20			
Mean Fit Time = 121.3s		Mean Fit Time = 177.8s		Mean Fit Time = 236.4s		Mean Fit Time = 295.8s			
RMSE		RMSE		RMSE		RMSE			
CROSS-VALIDATION (K = 10)	SPLIT 1	40.1473	SPLIT 1	39.2285	SPLIT 1	39.1817	SPLIT 1	39.2795	
	SPLIT 2	37.7154	SPLIT 2	37.3718	SPLIT 2	37.0715	SPLIT 2	37.1008	
	SPLIT 3	37.5586	SPLIT 3	37.7817	SPLIT 3	37.7453	SPLIT 3	37.2720	
	SPLIT 4	39.5153	SPLIT 4	39.3391	SPLIT 4	39.1627	SPLIT 4	39.2268	
	SPLIT 5	39.5487	SPLIT 5	39.7371	SPLIT 5	39.4850	SPLIT 5	39.2364	
	SPLIT 6	39.4639	SPLIT 6	39.2355	SPLIT 6	38.9547	SPLIT 6	39.1086	
	SPLIT 7	40.9932	SPLIT 7	40.4558	SPLIT 7	40.3418	SPLIT 7	40.4869	
	SPLIT 8	38.9475	SPLIT 8	39.0587	SPLIT 8	38.7690	SPLIT 8	38.7294	
	SPLIT 9	39.7266	SPLIT 9	39.9876	SPLIT 9	39.8030	SPLIT 9	39.3906	
	SPLIT 10	37.7668	SPLIT 10	37.5955	SPLIT 10	37.1868	SPLIT 10	37.1265	
MEAN	39.1383	MEAN	38.9791	MEAN	38.7701	MEAN	38.6957		
STD. DV.	1.0797	STD. DV.	0.9983	STD. DV.	1.0415	STD. DV.	1.0868		
MEDIAN	39.4896	MEDIAN	39.2320	MEDIAN	39.0587	MEDIAN	39.1677		
MIN.	37.5586	MIN.	37.3718	MIN.	37.0715	MIN.	37.1008		
MAX.	40.9932	MAX.	40.4558	MAX.	40.3418	MAX.	40.4869		

RANDOM FOREST									
<i>n_estimators</i> = 100 <i>max_depth</i> = 30		<i>n_estimators</i> = 150 <i>max_depth</i> = 30		<i>n_estimators</i> = 200 <i>max_depth</i> = 30		<i>n_estimators</i> = 250 <i>max_depth</i> = 30			
Mean Fit Time = 131.4s		Mean Fit Time = 197.3s		Mean Fit Time = 264.9s		Mean Fit Time = 330.4s			
RMSE		RMSE		RMSE		RMSE			
CROSS-VALIDATION (K = 10)	SPLIT 1	36.9613	SPLIT 1	36.7517	SPLIT 1	36.8718	SPLIT 1	36.8402	
	SPLIT 2	34.9547	SPLIT 2	34.7443	SPLIT 2	34.8124	SPLIT 2	34.6609	
	SPLIT 3	34.8713	SPLIT 3	34.9989	SPLIT 3	34.7051	SPLIT 3	34.9273	
	SPLIT 4	36.1380	SPLIT 4	36.3687	SPLIT 4	36.2188	SPLIT 4	36.2472	
	SPLIT 5	36.5316	SPLIT 5	36.9670	SPLIT 5	36.6295	SPLIT 5	36.7592	
	SPLIT 6	37.1930	SPLIT 6	36.7043	SPLIT 6	36.6142	SPLIT 6	36.6794	
	SPLIT 7	37.7502	SPLIT 7	37.5613	SPLIT 7	37.6799	SPLIT 7	37.8855	
	SPLIT 8	36.7986	SPLIT 8	36.5412	SPLIT 8	36.3186	SPLIT 8	36.5193	
	SPLIT 9	37.3247	SPLIT 9	37.6557	SPLIT 9	37.0203	SPLIT 9	36.9314	
	SPLIT 10	35.2758	SPLIT 10	34.7298	SPLIT 10	35.0535	SPLIT 10	34.6947	
MEAN	36.3799	MEAN	36.3023	MEAN	36.1924	MEAN	36.2145		
STD. DV.	0.9773	STD. DV.	1.0431	STD. DV.	0.9570	STD. DV.	1.0336		
MEDIAN	36.6651	MEDIAN	36.6227	MEDIAN	36.4664	MEDIAN	36.5993		
MIN.	34.8713	MIN.	34.7298	MIN.	34.7051	MIN.	34.6609		
MAX.	37.7502	MAX.	37.6557	MAX.	37.6799	MAX.	37.8855		

<b>XGBOOST</b>											
<i>learning_rate</i> = 0.001 <i>max_depth</i> = None <i>n_estimators</i> = 100		<i>learning_rate</i> = 0.001 <i>max_depth</i> = None <i>n_estimators</i> = 150		<i>learning_rate</i> = 0.001 <i>max_depth</i> = None <i>n_estimators</i> = 200		<i>learning_rate</i> = 0.001 <i>max_depth</i> = None <i>n_estimators</i> = 250					
<b>Mean Fit Time = 9.36s</b>		<b>Mean Fit Time = 14.1s</b>		<b>Mean Fit Time = 18.6s</b>		<b>Mean Fit Time = 23.2s</b>					
<b>RMSE</b>		<b>RMSE</b>		<b>RMSE</b>		<b>RMSE</b>					
<b>CROSS-VALIDATION (K = 10)</b>	<b>SPLIT 1</b>	1040.2311	<b>CROSS-VALIDATION (K = 10)</b>	<b>SPLIT 1</b>	993.1385	<b>CROSS-VALIDATION (K = 10)</b>	<b>SPLIT 1</b>	948.5300	<b>CROSS-VALIDATION (K = 10)</b>	<b>SPLIT 1</b>	906.2688
	<b>SPLIT 2</b>	1032.4914		<b>SPLIT 2</b>	985.6534		<b>SPLIT 2</b>	941.1961		<b>SPLIT 2</b>	899.0107
	<b>SPLIT 3</b>	1035.0164		<b>SPLIT 3</b>	988.1312		<b>SPLIT 3</b>	943.6212		<b>SPLIT 3</b>	901.4273
	<b>SPLIT 4</b>	1037.6035		<b>SPLIT 4</b>	990.6163		<b>SPLIT 4</b>	945.9661		<b>SPLIT 4</b>	903.6546
	<b>SPLIT 5</b>	1035.6476		<b>SPLIT 5</b>	988.8299		<b>SPLIT 5</b>	944.4482		<b>SPLIT 5</b>	902.3706
	<b>SPLIT 6</b>	1039.7621		<b>SPLIT 6</b>	992.7302		<b>SPLIT 6</b>	948.0959		<b>SPLIT 6</b>	905.7322
	<b>SPLIT 7</b>	1033.1290		<b>SPLIT 7</b>	986.2746		<b>SPLIT 7</b>	941.7660		<b>SPLIT 7</b>	899.5293
	<b>SPLIT 8</b>	1029.6254		<b>SPLIT 8</b>	982.8613		<b>SPLIT 8</b>	938.5293		<b>SPLIT 8</b>	896.4467
	<b>SPLIT 9</b>	1032.3948		<b>SPLIT 9</b>	985.4639		<b>SPLIT 9</b>	940.9729		<b>SPLIT 9</b>	898.7675
	<b>SPLIT 10</b>	1034.3311		<b>SPLIT 10</b>	987.4447		<b>SPLIT 10</b>	942.9646		<b>SPLIT 10</b>	900.7146
<b>MEAN</b>	1035.0232	<b>MEAN</b>	988.1144	<b>MEAN</b>	943.6090	<b>MEAN</b>	901.3922				
<b>STD. DV.</b>	3.2117	<b>STD. DV.</b>	3.1278	<b>STD. DV.</b>	3.0495	<b>STD. DV.</b>	2.9940				
<b>MEDIAN</b>	1034.6738	<b>MEDIAN</b>	987.7880	<b>MEDIAN</b>	943.2929	<b>MEDIAN</b>	901.0709				
<b>MIN.</b>	1029.6254	<b>MIN.</b>	982.8613	<b>MIN.</b>	938.5293	<b>MIN.</b>	896.4467				
<b>MAX.</b>	1040.2311	<b>MAX.</b>	993.1385	<b>MAX.</b>	948.5300	<b>MAX.</b>	906.2688				

<b>XGBOOST</b>								
<i>learning_rate</i> = 0.001 <i>max_depth</i> = 10 <i>n_estimators</i> = 100		<i>learning_rate</i> = 0.001 <i>max_depth</i> = 10 <i>n_estimators</i> = 150		<i>learning_rate</i> = 0.001 <i>max_depth</i> = 10 <i>n_estimators</i> = 200		<i>learning_rate</i> = 0.001 <i>max_depth</i> = 10 <i>n_estimators</i> = 250		
<b>Mean Fit Time = 16.1s</b>		<b>Mean Fit Time = 24.2s</b>		<b>Mean Fit Time = 32.2s</b>		<b>Mean Fit Time = 40.3s</b>		
<b>RMSE</b>		<b>RMSE</b>		<b>RMSE</b>		<b>RMSE</b>		
<b>CROSS-VALIDATION (K = 10)</b>	<b>SPLIT 1</b>	1036.9661	<b>SPLIT 1</b>	988.2475	<b>SPLIT 1</b>	941.9495	<b>SPLIT 1</b>	897.9545
	<b>SPLIT 2</b>	1029.3149	<b>SPLIT 2</b>	980.7965	<b>SPLIT 2</b>	934.7596	<b>SPLIT 2</b>	890.9891
	<b>SPLIT 3</b>	1031.5438	<b>SPLIT 3</b>	982.8909	<b>SPLIT 3</b>	936.6532	<b>SPLIT 3</b>	892.6908
	<b>SPLIT 4</b>	1034.1920	<b>SPLIT 4</b>	985.4826	<b>SPLIT 4</b>	939.1744	<b>SPLIT 4</b>	895.1463
	<b>SPLIT 5</b>	1032.1462	<b>SPLIT 5</b>	983.4849	<b>SPLIT 5</b>	937.2118	<b>SPLIT 5</b>	893.2534
	<b>SPLIT 6</b>	1036.2737	<b>SPLIT 6</b>	987.4447	<b>SPLIT 6</b>	941.0910	<b>SPLIT 6</b>	897.0561
	<b>SPLIT 7</b>	1029.8531	<b>SPLIT 7</b>	981.3294	<b>SPLIT 7</b>	935.2179	<b>SPLIT 7</b>	891.3894
	<b>SPLIT 8</b>	1026.3249	<b>SPLIT 8</b>	977.8868	<b>SPLIT 8</b>	931.8284	<b>SPLIT 8</b>	888.0498
	<b>SPLIT 9</b>	1029.2357	<b>SPLIT 9</b>	980.6627	<b>SPLIT 9</b>	934.5550	<b>SPLIT 9</b>	890.7513
	<b>SPLIT 10</b>	1031.0429	<b>SPLIT 10</b>	982.4778	<b>SPLIT 10</b>	936.3332	<b>SPLIT 10</b>	892.4898
<b>MEAN</b>	1031.6893	<b>MEAN</b>	983.0704	<b>MEAN</b>	936.8774	<b>MEAN</b>	892.9770	
<b>STD. DV.</b>	3.1556	<b>STD. DV.</b>	3.0525	<b>STD. DV.</b>	2.9553	<b>STD. DV.</b>	2.8673	
<b>MEDIAN</b>	1031.2933	<b>MEDIAN</b>	982.6843	<b>MEDIAN</b>	936.4932	<b>MEDIAN</b>	892.5903	
<b>MIN.</b>	1026.3249	<b>MIN.</b>	977.8868	<b>MIN.</b>	931.8284	<b>MIN.</b>	888.0498	
<b>MAX.</b>	1036.9661	<b>MAX.</b>	988.2475	<b>MAX.</b>	941.9495	<b>MAX.</b>	897.9545	

<b>XGBOOST</b>								
<i>learning_rate</i> = 0.001 <i>max_depth</i> = 20 <i>n_estimators</i> = 100		<i>learning_rate</i> = 0.001 <i>max_depth</i> = 20 <i>n_estimators</i> = 150		<i>learning_rate</i> = 0.001 <i>max_depth</i> = 20 <i>n_estimators</i> = 200		<i>learning_rate</i> = 0.001 <i>max_depth</i> = 20 <i>n_estimators</i> = 250		
<b>Mean Fit Time = 28.1s</b>		<b>Mean Fit Time = 42.1s</b>		<b>Mean Fit Time = 56.3s</b>		<b>Mean Fit Time = 70.9s</b>		
<b>RMSE</b>		<b>RMSE</b>		<b>RMSE</b>		<b>RMSE</b>		
<b>CROSS-VALIDATION (K = 10)</b>	<b>SPLIT 1</b>	1034.5987	<b>SPLIT 1</b>	984.6770	<b>SPLIT 1</b>	937.1923	<b>SPLIT 1</b>	892.0222
	<b>SPLIT 2</b>	1027.0991	<b>SPLIT 2</b>	977.5136	<b>SPLIT 2</b>	930.3392	<b>SPLIT 2</b>	885.4647
	<b>SPLIT 3</b>	1029.4839	<b>SPLIT 3</b>	979.8115	<b>SPLIT 3</b>	932.5779	<b>SPLIT 3</b>	887.6497
	<b>SPLIT 4</b>	1032.0981	<b>SPLIT 4</b>	982.3166	<b>SPLIT 4</b>	934.9643	<b>SPLIT 4</b>	889.9116
	<b>SPLIT 5</b>	1029.9557	<b>SPLIT 5</b>	980.2676	<b>SPLIT 5</b>	933.0030	<b>SPLIT 5</b>	888.0389
	<b>SPLIT 6</b>	1034.0978	<b>SPLIT 6</b>	984.1930	<b>SPLIT 6</b>	936.7182	<b>SPLIT 6</b>	891.5694
	<b>SPLIT 7</b>	1027.7271	<b>SPLIT 7</b>	978.1356	<b>SPLIT 7</b>	930.9709	<b>SPLIT 7</b>	886.0940
	<b>SPLIT 8</b>	1024.4858	<b>SPLIT 8</b>	975.1193	<b>SPLIT 8</b>	928.1367	<b>SPLIT 8</b>	883.4371
	<b>SPLIT 9</b>	1027.1363	<b>SPLIT 9</b>	977.5179	<b>SPLIT 9</b>	930.3311	<b>SPLIT 9</b>	885.4408
	<b>SPLIT 10</b>	1028.8676	<b>SPLIT 10</b>	979.1965	<b>SPLIT 10</b>	931.9427	<b>SPLIT 10</b>	886.9912
<b>MEAN</b>	1029.5550	<b>MEAN</b>	979.8749	<b>MEAN</b>	932.6176	<b>MEAN</b>	887.6620	
<b>STD. DV.</b>	3.0675	<b>STD. DV.</b>	2.9164	<b>STD. DV.</b>	2.7760	<b>STD. DV.</b>	2.6458	
<b>MEDIAN</b>	1029.1757	<b>MEDIAN</b>	979.5040	<b>MEDIAN</b>	932.2603	<b>MEDIAN</b>	887.3204	
<b>MIN.</b>	1024.4858	<b>MIN.</b>	975.1193	<b>MIN.</b>	928.1367	<b>MIN.</b>	883.4371	
<b>MAX.</b>	1034.5987	<b>MAX.</b>	984.6770	<b>MAX.</b>	937.1923	<b>MAX.</b>	892.0222	

<b>XGBOOST</b>											
<i>learning_rate</i> = 0.001 <i>max_depth</i> = 30 <i>n_estimators</i> = 100		<i>learning_rate</i> = 0.001 <i>max_depth</i> = 30 <i>n_estimators</i> = 150		<i>learning_rate</i> = 0.001 <i>max_depth</i> = 30 <i>n_estimators</i> = 200		<i>learning_rate</i> = 0.001 <i>max_depth</i> = 30 <i>n_estimators</i> = 250					
<b>Mean Fit Time = 29.3s</b>		<b>Mean Fit Time = 44.1s</b>		<b>Mean Fit Time = 59.0s</b>		<b>Mean Fit Time = 74.6s</b>					
<b>RMSE</b>		<b>RMSE</b>		<b>RMSE</b>		<b>RMSE</b>					
<b>CROSS-VALIDATION (K = 10)</b>	<b>SPLIT 1</b>	1034.5873	<b>CROSS-VALIDATION (K = 10)</b>	<b>SPLIT 1</b>	984.6613	<b>CROSS-VALIDATION (K = 10)</b>	<b>SPLIT 1</b>	937.1695	<b>CROSS-VALIDATION (K = 10)</b>	<b>SPLIT 1</b>	891.9941
	<b>SPLIT 2</b>	1027.0958		<b>SPLIT 2</b>	977.5080		<b>SPLIT 2</b>	930.3343		<b>SPLIT 2</b>	885.4581
	<b>SPLIT 3</b>	1029.4630		<b>SPLIT 3</b>	979.7752		<b>SPLIT 3</b>	932.5179		<b>SPLIT 3</b>	887.5714
	<b>SPLIT 4</b>	1032.0789		<b>SPLIT 4</b>	982.2842		<b>SPLIT 4</b>	934.9152		<b>SPLIT 4</b>	889.8554
	<b>SPLIT 5</b>	1029.9412		<b>SPLIT 5</b>	980.2436		<b>SPLIT 5</b>	932.9688		<b>SPLIT 5</b>	887.9966
	<b>SPLIT 6</b>	1034.0874		<b>SPLIT 6</b>	984.1762		<b>SPLIT 6</b>	936.6937		<b>SPLIT 6</b>	891.5368
	<b>SPLIT 7</b>	1027.7181		<b>SPLIT 7</b>	978.1260		<b>SPLIT 7</b>	930.9602		<b>SPLIT 7</b>	886.0842
	<b>SPLIT 8</b>	1024.4775		<b>SPLIT 8</b>	975.1100		<b>SPLIT 8</b>	928.1240		<b>SPLIT 8</b>	883.4227
	<b>SPLIT 9</b>	1027.1313		<b>SPLIT 9</b>	977.5127		<b>SPLIT 9</b>	930.3236		<b>SPLIT 9</b>	885.4305
	<b>SPLIT 10</b>	1028.8652		<b>SPLIT 10</b>	979.1938		<b>SPLIT 10</b>	931.9393		<b>SPLIT 10</b>	886.9870
<b>MEAN</b>	1029.5446	<b>MEAN</b>	979.8591	<b>MEAN</b>	932.5947	<b>MEAN</b>	887.6337				
<b>STD. DV.</b>	3.0650	<b>STD. DV.</b>	2.9114	<b>STD. DV.</b>	2.7679	<b>STD. DV.</b>	2.6354				
<b>MEDIAN</b>	1029.1641	<b>MEDIAN</b>	979.4845	<b>MEDIAN</b>	932.2286	<b>MEDIAN</b>	887.2792				
<b>MIN.</b>	1024.4775	<b>MIN.</b>	975.1100	<b>MIN.</b>	928.1240	<b>MIN.</b>	883.4227				
<b>MAX.</b>	1034.5873	<b>MAX.</b>	984.6613	<b>MAX.</b>	937.1695	<b>MAX.</b>	891.9941				

<b>XGBOOST</b>								
<i>learning_rate</i> = 0.01 <i>max_depth</i> = None <i>n_estimators</i> = 100		<i>learning_rate</i> = 0.01 <i>max_depth</i> = None <i>n_estimators</i> = 150		<i>learning_rate</i> = 0.01 <i>max_depth</i> = None <i>n_estimators</i> = 200		<i>learning_rate</i> = 0.01 <i>max_depth</i> = None <i>n_estimators</i> = 250		
<b>Mean Fit Time = 9.45s</b>		<b>Mean Fit Time = 14.1s</b>		<b>Mean Fit Time = 18.7s</b>		<b>Mean Fit Time = 23.6s</b>		
<b>RMSE</b>		<b>RMSE</b>		<b>RMSE</b>		<b>RMSE</b>		
<b>CROSS-VALIDATION (K = 10)</b>	<b>SPLIT 1</b>	483.9001	<b>CROSS-VALIDATION (K = 10)</b>	<b>SPLIT 1</b>	349.7236	<b>CROSS-VALIDATION (K = 10)</b>	<b>SPLIT 1</b>	240.2260
	<b>SPLIT 2</b>	477.9117		<b>SPLIT 2</b>	344.5434		<b>SPLIT 2</b>	235.5334
	<b>SPLIT 3</b>	480.4292		<b>SPLIT 3</b>	346.9236		<b>SPLIT 3</b>	238.0455
	<b>SPLIT 4</b>	481.1641		<b>SPLIT 4</b>	347.1802		<b>SPLIT 4</b>	237.7889
	<b>SPLIT 5</b>	481.3927		<b>SPLIT 5</b>	347.7398		<b>SPLIT 5</b>	238.6411
	<b>SPLIT 6</b>	482.6616		<b>SPLIT 6</b>	347.9425		<b>SPLIT 6</b>	237.8317
	<b>SPLIT 7</b>	478.4927		<b>SPLIT 7</b>	345.0305		<b>SPLIT 7</b>	236.9101
	<b>SPLIT 8</b>	476.1986		<b>SPLIT 8</b>	343.2241		<b>SPLIT 8</b>	235.3571
	<b>SPLIT 9</b>	477.8227		<b>SPLIT 9</b>	344.4274		<b>SPLIT 9</b>	236.1971
	<b>SPLIT 10</b>	478.7034		<b>SPLIT 10</b>	344.7911		<b>SPLIT 10</b>	236.0570
<b>MEAN</b>	479.8677	<b>MEAN</b>	346.1526	<b>MEAN</b>	273.9967	<b>MEAN</b>	237.2588	
<b>STD. DV.</b>	2.3044	<b>STD. DV.</b>	1.9334	<b>STD. DV.</b>	1.7547	<b>STD. DV.</b>	1.4570	
<b>MEDIAN</b>	479.5663	<b>MEDIAN</b>	345.9770	<b>MEDIAN</b>	274.0025	<b>MEDIAN</b>	237.3495	
<b>MIN.</b>	476.1986	<b>MIN.</b>	343.2241	<b>MIN.</b>	271.4134	<b>MIN.</b>	235.3571	
<b>MAX.</b>	483.9001	<b>MAX.</b>	349.7236	<b>MAX.</b>	277.3903	<b>MAX.</b>	240.2260	

<b>XGBOOST</b>											
<i>learning_rate</i> = 0.01 <i>max_depth</i> = 10 <i>n_estimators</i> = 100		<i>learning_rate</i> = 0.01 <i>max_depth</i> = 10 <i>n_estimators</i> = 150		<i>learning_rate</i> = 0.01 <i>max_depth</i> = 10 <i>n_estimators</i> = 200		<i>learning_rate</i> = 0.01 <i>max_depth</i> = 10 <i>n_estimators</i> = 250					
<b>Mean Fit Time = 16.3s</b>		<b>Mean Fit Time = 24.6s</b>		<b>Mean Fit Time = 32.9s</b>		<b>Mean Fit Time = 41.4s</b>					
<b>RMSE</b>		<b>RMSE</b>		<b>RMSE</b>		<b>RMSE</b>					
<b>CROSS-VALIDATION (K = 10)</b>	<b>SPLIT 1</b>	446.2724	<b>CROSS-VALIDATION (K = 10)</b>	<b>SPLIT 1</b>	291.6975	<b>CROSS-VALIDATION (K = 10)</b>	<b>SPLIT 1</b>	202.2329	<b>CROSS-VALIDATION (K = 10)</b>	<b>SPLIT 1</b>	149.8897
	<b>SPLIT 2</b>	441.4905		<b>SPLIT 2</b>	288.2173		<b>SPLIT 2</b>	199.1258		<b>SPLIT 2</b>	148.5241
	<b>SPLIT 3</b>	442.4031		<b>SPLIT 3</b>	289.1253		<b>SPLIT 3</b>	199.5709		<b>SPLIT 3</b>	148.1198
	<b>SPLIT 4</b>	444.4568		<b>SPLIT 4</b>	290.4188		<b>SPLIT 4</b>	200.5644		<b>SPLIT 4</b>	148.6789
	<b>SPLIT 5</b>	443.5795		<b>SPLIT 5</b>	290.6130		<b>SPLIT 5</b>	201.9046		<b>SPLIT 5</b>	150.4814
	<b>SPLIT 6</b>	445.7181		<b>SPLIT 6</b>	291.3681		<b>SPLIT 6</b>	201.0607		<b>SPLIT 6</b>	148.9169
	<b>SPLIT 7</b>	441.9814		<b>SPLIT 7</b>	288.5774		<b>SPLIT 7</b>	198.6813		<b>SPLIT 7</b>	147.9916
	<b>SPLIT 8</b>	439.2530		<b>SPLIT 8</b>	286.7811		<b>SPLIT 8</b>	198.5109		<b>SPLIT 8</b>	148.2816
	<b>SPLIT 9</b>	441.4799		<b>SPLIT 9</b>	288.2343		<b>SPLIT 9</b>	199.5674		<b>SPLIT 9</b>	148.2991
	<b>SPLIT 10</b>	442.2523		<b>SPLIT 10</b>	288.1870		<b>SPLIT 10</b>	198.7877		<b>SPLIT 10</b>	148.1027
<b>MEAN</b>	442.8887	<b>MEAN</b>	289.3220	<b>MEAN</b>	200.0007	<b>MEAN</b>	148.7286				
<b>STD. DV.</b>	2.0251	<b>STD. DV.</b>	1.5310	<b>STD. DV.</b>	1.2891	<b>STD. DV.</b>	0.7867				
<b>MEDIAN</b>	442.3277	<b>MEDIAN</b>	288.85135	<b>MEDIAN</b>	199.5691	<b>MEDIAN</b>	148.4116				
<b>MIN.</b>	439.2530	<b>MIN.</b>	286.7811	<b>MIN.</b>	198.5109	<b>MIN.</b>	147.9916				
<b>MAX.</b>	446.2724	<b>MAX.</b>	291.6975	<b>MAX.</b>	202.2329	<b>MAX.</b>	150.4814				

<b>XGBOOST</b>											
<i>learning_rate</i> = 0.01 <i>max_depth</i> = 20 <i>n_estimators</i> = 100		<i>learning_rate</i> = 0.01 <i>max_depth</i> = 20 <i>n_estimators</i> = 150		<i>learning_rate</i> = 0.01 <i>max_depth</i> = 20 <i>n_estimators</i> = 200		<i>learning_rate</i> = 0.01 <i>max_depth</i> = 20 <i>n_estimators</i> = 250					
<b>Mean Fit Time = 29.1s</b>		<b>Mean Fit Time = 44.3s</b>		<b>Mean Fit Time = 60.6s</b>		<b>Mean Fit Time = 78.3s</b>					
<b>RMSE</b>		<b>RMSE</b>		<b>RMSE</b>		<b>RMSE</b>					
<b>CROSS-VALIDATION (K = 10)</b>	<b>SPLIT 1</b>	425.0234	<b>CROSS-VALIDATION (K = 10)</b>	<b>SPLIT 1</b>	261.4089	<b>CROSS-VALIDATION (K = 10)</b>	<b>SPLIT 1</b>	163.1710	<b>CROSS-VALIDATION (K = 10)</b>	<b>SPLIT 1</b>	104.9389
	<b>SPLIT 2</b>	421.5832		<b>SPLIT 2</b>	259.0031		<b>SPLIT 2</b>	161.2823		<b>SPLIT 2</b>	103.2446
	<b>SPLIT 3</b>	422.8096		<b>SPLIT 3</b>	259.9616		<b>SPLIT 3</b>	162.0052		<b>SPLIT 3</b>	103.9705
	<b>SPLIT 4</b>	424.0495		<b>SPLIT 4</b>	260.9643		<b>SPLIT 4</b>	162.8772		<b>SPLIT 4</b>	104.6553
	<b>SPLIT 5</b>	423.2694		<b>SPLIT 5</b>	260.8896		<b>SPLIT 5</b>	163.3867		<b>SPLIT 5</b>	105.7954
	<b>SPLIT 6</b>	424.9475		<b>SPLIT 6</b>	261.5896		<b>SPLIT 6</b>	163.4511		<b>SPLIT 6</b>	105.3045
	<b>SPLIT 7</b>	421.9976		<b>SPLIT 7</b>	259.6235		<b>SPLIT 7</b>	162.2569		<b>SPLIT 7</b>	104.5599
	<b>SPLIT 8</b>	421.1706		<b>SPLIT 8</b>	259.2862		<b>SPLIT 8</b>	162.1933		<b>SPLIT 8</b>	104.6942
	<b>SPLIT 9</b>	421.5358		<b>SPLIT 9</b>	259.0518		<b>SPLIT 9</b>	161.3935		<b>SPLIT 9</b>	103.6515
	<b>SPLIT 10</b>	422.1758		<b>SPLIT 10</b>	259.3442		<b>SPLIT 10</b>	161.4590		<b>SPLIT 10</b>	103.4259
<b>MEAN</b>	422.8562	<b>MEAN</b>	260.1123	<b>MEAN</b>	162.3476	<b>MEAN</b>	104.4241				
<b>STD. DV.</b>	1.3443	<b>STD. DV.</b>	0.9530	<b>STD. DV.</b>	0.7899	<b>STD. DV.</b>	0.7909				
<b>MEDIAN</b>	422.4927	<b>MEDIAN</b>	259.7925	<b>MEDIAN</b>	162.2251	<b>MEDIAN</b>	104.6076				
<b>MIN.</b>	421.1706	<b>MIN.</b>	259.0031	<b>MIN.</b>	161.2823	<b>MIN.</b>	103.2446				
<b>MAX.</b>	425.0234	<b>MAX.</b>	261.5896	<b>MAX.</b>	163.4511	<b>MAX.</b>	105.7954				

<b>XGBOOST</b>											
<i>learning_rate</i> = 0.01 <i>max_depth</i> = 30 <i>n_estimators</i> = 100		<i>learning_rate</i> = 0.01 <i>max_depth</i> = 30 <i>n_estimators</i> = 150		<i>learning_rate</i> = 0.01 <i>max_depth</i> = 30 <i>n_estimators</i> = 200		<i>learning_rate</i> = 0.01 <i>max_depth</i> = 30 <i>n_estimators</i> = 250					
<b>Mean Fit Time = 31.0s</b>		<b>Mean Fit Time = 47.9s</b>		<b>Mean Fit Time = 66.6s</b>		<b>Mean Fit Time = 87.8s</b>					
<b>RMSE</b>		<b>RMSE</b>		<b>RMSE</b>		<b>RMSE</b>					
<b>CROSS-VALIDATION (K = 10)</b>	<b>SPLIT 1</b>	424.9546	<b>CROSS-VALIDATION (K = 10)</b>	<b>SPLIT 1</b>	261.3485	<b>CROSS-VALIDATION (K = 10)</b>	<b>SPLIT 1</b>	163.1186	<b>CROSS-VALIDATION (K = 10)</b>	<b>SPLIT 1</b>	105.0023
	<b>SPLIT 2</b>	421.5559		<b>SPLIT 2</b>	258.9926		<b>SPLIT 2</b>	161.2071		<b>SPLIT 2</b>	103.1836
	<b>SPLIT 3</b>	422.6204		<b>SPLIT 3</b>	259.8351		<b>SPLIT 3</b>	161.8977		<b>SPLIT 3</b>	103.9304
	<b>SPLIT 4</b>	423.8903		<b>SPLIT 4</b>	260.7772		<b>SPLIT 4</b>	162.7113		<b>SPLIT 4</b>	104.4498
	<b>SPLIT 5</b>	423.1896		<b>SPLIT 5</b>	260.8379		<b>SPLIT 5</b>	163.3740		<b>SPLIT 5</b>	105.7820
	<b>SPLIT 6</b>	424.9092		<b>SPLIT 6</b>	261.5898		<b>SPLIT 6</b>	163.4289		<b>SPLIT 6</b>	105.3314
	<b>SPLIT 7</b>	422.0176		<b>SPLIT 7</b>	259.5563		<b>SPLIT 7</b>	162.0655		<b>SPLIT 7</b>	104.3917
	<b>SPLIT 8</b>	421.0864		<b>SPLIT 8</b>	259.2042		<b>SPLIT 8</b>	162.0859		<b>SPLIT 8</b>	104.6790
	<b>SPLIT 9</b>	421.5074		<b>SPLIT 9</b>	259.0239		<b>SPLIT 9</b>	161.3643		<b>SPLIT 9</b>	103.6934
	<b>SPLIT 10</b>	422.0771		<b>SPLIT 10</b>	259.1911		<b>SPLIT 10</b>	161.2613		<b>SPLIT 10</b>	103.1768
<b>MEAN</b>	422.7809	<b>MEAN</b>	260.0357	<b>MEAN</b>	162.2515	<b>MEAN</b>	104.3620				
<b>STD. DV.</b>	1.3322	<b>STD. DV.</b>	0.9549	<b>STD. DV.</b>	0.8157	<b>STD. DV.</b>	0.8307				
<b>MEDIAN</b>	422.3487	<b>MEDIAN</b>	259.6957	<b>MEDIAN</b>	162.0757	<b>MEDIAN</b>	104.4207				
<b>MIN.</b>	421.0864	<b>MIN.</b>	258.9926	<b>MIN.</b>	161.2071	<b>MIN.</b>	103.1768				
<b>MAX.</b>	424.9546	<b>MAX.</b>	261.5898	<b>MAX.</b>	163.4289	<b>MAX.</b>	105.7820				

XGBOOST								
<i>learning_rate</i> = 0.1 <i>max_depth</i> = None <i>n_estimators</i> = 100		<i>learning_rate</i> = 0.1 <i>max_depth</i> = None <i>n_estimators</i> = 150		<i>learning_rate</i> = 0.1 <i>max_depth</i> = None <i>n_estimators</i> = 200		<i>learning_rate</i> = 0.1 <i>max_depth</i> = None <i>n_estimators</i> = 250		
<b>Mean Fit Time = 9.54s</b>		<b>Mean Fit Time = 14.3s</b>		<b>Mean Fit Time = 19.1s</b>		<b>Mean Fit Time = 23.8s</b>		
<b>RMSE</b>		<b>RMSE</b>		<b>RMSE</b>		<b>RMSE</b>		
<b>CROSS-VALIDATION (K = 10)</b>	<b>SPLIT 1</b>	150.6273	<b>SPLIT 1</b>	126.1758	<b>SPLIT 1</b>	110.0714	<b>SPLIT 1</b>	97.5930
	<b>SPLIT 2</b>	151.2363	<b>SPLIT 2</b>	125.1317	<b>SPLIT 2</b>	110.2132	<b>SPLIT 2</b>	98.9567
	<b>SPLIT 3</b>	151.6538	<b>SPLIT 3</b>	128.2919	<b>SPLIT 3</b>	110.6212	<b>SPLIT 3</b>	96.4959
	<b>SPLIT 4</b>	153.5737	<b>SPLIT 4</b>	127.0455	<b>SPLIT 4</b>	113.0583	<b>SPLIT 4</b>	100.6009
	<b>SPLIT 5</b>	156.8830	<b>SPLIT 5</b>	130.5391	<b>SPLIT 5</b>	113.6050	<b>SPLIT 5</b>	99.6231
	<b>SPLIT 6</b>	152.0525	<b>SPLIT 6</b>	127.8402	<b>SPLIT 6</b>	110.0600	<b>SPLIT 6</b>	96.7843
	<b>SPLIT 7</b>	151.1583	<b>SPLIT 7</b>	130.2408	<b>SPLIT 7</b>	113.9029	<b>SPLIT 7</b>	99.6465
	<b>SPLIT 8</b>	153.5530	<b>SPLIT 8</b>	129.1028	<b>SPLIT 8</b>	112.9741	<b>SPLIT 8</b>	100.0576
	<b>SPLIT 9</b>	154.1818	<b>SPLIT 9</b>	126.7844	<b>SPLIT 9</b>	110.9070	<b>SPLIT 9</b>	99.5120
	<b>SPLIT 10</b>	150.9936	<b>SPLIT 10</b>	123.9384	<b>SPLIT 10</b>	108.6455	<b>SPLIT 10</b>	97.6199
<b>MEAN</b>	152.5913	<b>MEAN</b>	127.5090	<b>MEAN</b>	111.4059	<b>MEAN</b>	98.6890	
<b>STD. DV.</b>	1.8522	<b>STD. DV.</b>	2.0205	<b>STD. DV.</b>	1.7251	<b>STD. DV.</b>	1.3728	
<b>MEDIAN</b>	151.8531	<b>MEDIAN</b>	127.4428	<b>MEDIAN</b>	110.7641	<b>MEDIAN</b>	99.2343	
<b>MIN.</b>	150.6273	<b>MIN.</b>	123.9384	<b>MIN.</b>	108.6455	<b>MIN.</b>	96.4959	
<b>MAX.</b>	156.8830	<b>MAX.</b>	130.5391	<b>MAX.</b>	113.9029	<b>MAX.</b>	100.6009	

<b>XGBOOST</b>								
<i>learning_rate</i> = 0.1 <i>max_depth</i> = 10 <i>n_estimators</i> = 100		<i>learning_rate</i> = 0.1 <i>max_depth</i> = 10 <i>n_estimators</i> = 150		<i>learning_rate</i> = 0.1 <i>max_depth</i> = 10 <i>n_estimators</i> = 200		<i>learning_rate</i> = 0.1 <i>max_depth</i> = 10 <i>n_estimators</i> = 250		
<b>Mean Fit Time = 16.9s</b>		<b>Mean Fit Time = 25.5s</b>		<b>Mean Fit Time = 34.0s</b>		<b>Mean Fit Time = 42.6s</b>		
<b>RMSE</b>		<b>RMSE</b>		<b>RMSE</b>		<b>RMSE</b>		
<b>CROSS-VALIDATION (K = 10)</b>	<b>SPLIT 1</b>	57.9800	<b>SPLIT 1</b>	48.3334	<b>SPLIT 1</b>	44.0515	<b>SPLIT 1</b>	42.1792
	<b>SPLIT 2</b>	56.8135	<b>SPLIT 2</b>	48.5726	<b>SPLIT 2</b>	44.3182	<b>SPLIT 2</b>	42.0024
	<b>SPLIT 3</b>	59.4518	<b>SPLIT 3</b>	50.0918	<b>SPLIT 3</b>	45.1502	<b>SPLIT 3</b>	43.0361
	<b>SPLIT 4</b>	57.3506	<b>SPLIT 4</b>	47.7731	<b>SPLIT 4</b>	44.6152	<b>SPLIT 4</b>	42.4284
	<b>SPLIT 5</b>	59.6600	<b>SPLIT 5</b>	50.4870	<b>SPLIT 5</b>	45.9817	<b>SPLIT 5</b>	43.3585
	<b>SPLIT 6</b>	61.8298	<b>SPLIT 6</b>	49.5819	<b>SPLIT 6</b>	45.1833	<b>SPLIT 6</b>	42.6976
	<b>SPLIT 7</b>	57.4637	<b>SPLIT 7</b>	48.3288	<b>SPLIT 7</b>	44.9070	<b>SPLIT 7</b>	42.9368
	<b>SPLIT 8</b>	61.0649	<b>SPLIT 8</b>	50.9443	<b>SPLIT 8</b>	46.1812	<b>SPLIT 8</b>	43.9306
	<b>SPLIT 9</b>	59.2795	<b>SPLIT 9</b>	49.4208	<b>SPLIT 9</b>	44.7290	<b>SPLIT 9</b>	42.1540
	<b>SPLIT 10</b>	59.4872	<b>SPLIT 10</b>	48.4289	<b>SPLIT 10</b>	43.9710	<b>SPLIT 10</b>	42.2940
<b>MEAN</b>	59.0381	<b>MEAN</b>	49.1962	<b>MEAN</b>	44.9088	<b>MEAN</b>	42.7018	
<b>STD. DV.</b>	1.5528	<b>STD. DV.</b>	1.0110	<b>STD. DV.</b>	0.7060	<b>STD. DV.</b>	0.5840	
<b>MEDIAN</b>	59.3656	<b>MEDIAN</b>	48.9967	<b>MEDIAN</b>	44.8180	<b>MEDIAN</b>	42.5630	
<b>MIN.</b>	56.8135	<b>MIN.</b>	47.7731	<b>MIN.</b>	43.9710	<b>MIN.</b>	42.0024	
<b>MAX.</b>	61.8298	<b>MAX.</b>	50.9443	<b>MAX.</b>	46.1812	<b>MAX.</b>	43.9306	

<b>XGBOOST</b>								
<i>learning_rate</i> = 0.1 <i>max_depth</i> = 20 <i>n_estimators</i> = 100		<i>learning_rate</i> = 0.1 <i>max_depth</i> = 20 <i>n_estimators</i> = 150		<i>learning_rate</i> = 0.1 <i>max_depth</i> = 20 <i>n_estimators</i> = 200		<i>learning_rate</i> = 0.1 <i>max_depth</i> = 20 <i>n_estimators</i> = 250		
<b>Mean Fit Time = 43.9s</b>		<b>Mean Fit Time = 62.8s</b>		<b>Mean Fit Time = 81.4s</b>		<b>Mean Fit Time = 100.2s</b>		
<b>RMSE</b>		<b>RMSE</b>		<b>RMSE</b>		<b>RMSE</b>		
<b>CROSS-VALIDATION (K = 10)</b>	<b>SPLIT 1</b>	35.9692	<b>SPLIT 1</b>	35.8916	<b>SPLIT 1</b>	35.8536	<b>SPLIT 1</b>	35.8371
	<b>SPLIT 2</b>	33.4491	<b>SPLIT 2</b>	33.4010	<b>SPLIT 2</b>	33.3674	<b>SPLIT 2</b>	33.3228
	<b>SPLIT 3</b>	35.2595	<b>SPLIT 3</b>	35.2088	<b>SPLIT 3</b>	35.1673	<b>SPLIT 3</b>	35.1458
	<b>SPLIT 4</b>	35.4109	<b>SPLIT 4</b>	35.3567	<b>SPLIT 4</b>	35.2957	<b>SPLIT 4</b>	35.2858
	<b>SPLIT 5</b>	38.2610	<b>SPLIT 5</b>	38.1859	<b>SPLIT 5</b>	38.1690	<b>SPLIT 5</b>	38.1519
	<b>SPLIT 6</b>	36.7429	<b>SPLIT 6</b>	36.6938	<b>SPLIT 6</b>	36.6367	<b>SPLIT 6</b>	36.6088
	<b>SPLIT 7</b>	36.1512	<b>SPLIT 7</b>	36.0698	<b>SPLIT 7</b>	36.0315	<b>SPLIT 7</b>	36.0196
	<b>SPLIT 8</b>	37.7892	<b>SPLIT 8</b>	37.7103	<b>SPLIT 8</b>	37.6428	<b>SPLIT 8</b>	37.6320
	<b>SPLIT 9</b>	36.4956	<b>SPLIT 9</b>	36.4173	<b>SPLIT 9</b>	36.3855	<b>SPLIT 9</b>	36.3665
	<b>SPLIT 10</b>	34.4509	<b>SPLIT 10</b>	34.3762	<b>SPLIT 10</b>	34.3468	<b>SPLIT 10</b>	34.3256
<b>MEAN</b>	35.9980	<b>MEAN</b>	35.9311	<b>MEAN</b>	35.8896	<b>MEAN</b>	35.8696	
<b>STD. DV.</b>	1.3761	<b>STD. DV.</b>	1.3696	<b>STD. DV.</b>	1.3678	<b>STD. DV.</b>	1.3735	
<b>MEDIAN</b>	36.0602	<b>MEDIAN</b>	35.9807	<b>MEDIAN</b>	35.9425	<b>MEDIAN</b>	35.9283	
<b>MIN.</b>	33.4491	<b>MIN.</b>	33.4010	<b>MIN.</b>	33.3674	<b>MIN.</b>	33.3228	
<b>MAX.</b>	38.2610	<b>MAX.</b>	38.1859	<b>MAX.</b>	38.1690	<b>MAX.</b>	38.1519	

<b>XGBOOST</b>									
<i>learning_rate</i> = 0.1		<i>learning_rate</i> = 0.1		<i>learning_rate</i> = 0.1		<i>learning_rate</i> = 0.1			
<i>max_depth</i> = 30		<i>max_depth</i> = 30		<i>max_depth</i> = 30		<i>max_depth</i> = 30			
<i>n_estimators</i> = 100		<i>n_estimators</i> = 150		<i>n_estimators</i> = 200		<i>n_estimators</i> = 250			
<b>Mean Fit Time = 72.8s</b>		<b>Mean Fit Time = 119.9s</b>		<b>Mean Fit Time = 159.7s</b>		<b>Mean Fit Time = 193.9s</b>			
<b>RMSE</b>		<b>RMSE</b>		<b>RMSE</b>		<b>RMSE</b>			
<b>CROSS-VALIDATION (K = 10)</b>	<b>SPLIT 1</b>	35.8664	<b>SPLIT 1</b>	35.8412	<b>SPLIT 1</b>	35.8402	<b>SPLIT 1</b>	35.8401	
	<b>SPLIT 2</b>	33.8907	<b>SPLIT 2</b>	33.8737	<b>SPLIT 2</b>	33.8723	<b>SPLIT 2</b>	33.8723	
	<b>SPLIT 3</b>	35.5237	<b>SPLIT 3</b>	35.5008	<b>SPLIT 3</b>	35.4993	<b>SPLIT 3</b>	35.4993	
	<b>SPLIT 4</b>	35.8273	<b>SPLIT 4</b>	35.8116	<b>SPLIT 4</b>	35.8107	<b>SPLIT 4</b>	35.8106	
	<b>SPLIT 5</b>	38.4616	<b>SPLIT 5</b>	38.4406	<b>SPLIT 5</b>	38.4390	<b>SPLIT 5</b>	38.4390	
	<b>SPLIT 6</b>	37.5292	<b>SPLIT 6</b>	37.5090	<b>SPLIT 6</b>	37.5068	<b>SPLIT 6</b>	37.5064	
	<b>SPLIT 7</b>	36.8333	<b>SPLIT 7</b>	36.8213	<b>SPLIT 7</b>	36.8203	<b>SPLIT 7</b>	36.8203	
	<b>SPLIT 8</b>	37.6982	<b>SPLIT 8</b>	37.6812	<b>SPLIT 8</b>	37.6799	<b>SPLIT 8</b>	37.6799	
	<b>SPLIT 9</b>	36.8587	<b>SPLIT 9</b>	36.8450	<b>SPLIT 9</b>	36.8431	<b>SPLIT 9</b>	36.8430	
	<b>SPLIT 10</b>	34.3678	<b>SPLIT 10</b>	34.3536	<b>SPLIT 10</b>	34.3520	<b>SPLIT 10</b>	34.3518	
<b>MEAN</b>	36.2857	<b>MEAN</b>	36.2678	<b>MEAN</b>	36.2664	<b>MEAN</b>	36.2662		
<b>STD. DV.</b>	1.3887	<b>STD. DV.</b>	1.3882	<b>STD. DV.</b>	1.3881	<b>STD. DV.</b>	1.3881		
<b>MEDIAN</b>	36.3498	<b>MEDIAN</b>	36.3312	<b>MEDIAN</b>	36.3302	<b>MEDIAN</b>	36.3302		
<b>MIN.</b>	33.8907	<b>MIN.</b>	33.8737	<b>MIN.</b>	33.8723	<b>MIN.</b>	33.8723		
<b>MAX.</b>	38.4616	<b>MAX.</b>	38.4406	<b>MAX.</b>	38.4390	<b>MAX.</b>	38.4390		