

Análise de Séries Temporais de Queimadas no Brasil para a Gestão Ambiental e o Desenvolvimento Sustentável

Caroline Ribeiro Ferreira – 10408052,
Leonardo dos Reis Olher – 10407752,
Liliane Gonçalves de Brito Ferraz – 10407087,
Múcio Emanuel Feitosa Ferraz Filho – 10218691

¹Faculdade de Computação e Informática (FCI)
Universidade Presbiteriana Mackenzie – São Paulo, SP – Brasil

{gustavo.sampaio}@mackenzie.br

Resumo. Este estudo aprofunda a compreensão dos incêndios florestais no Brasil, utilizando técnicas avançadas de análise de dados. Através da análise de séries temporais de queimadas, o objetivo é desenvolver modelos preditivos capazes de antecipar a ocorrência desses eventos, auxiliando na gestão ambiental e na tomada de decisões. Ao identificar padrões, tendências e os principais fatores que influenciam as queimadas, a pesquisa contribui para a prevenção e o combate a incêndios, mitigando os impactos das mudanças climáticas e promovendo a sustentabilidade.

Palavras-chave: Brasil, queimadas, análise de dados, modelos preditivos, sustentabilidade.

Abstract. This study deepens the understanding of forest fires in Brazil using advanced data analysis techniques. Through the analysis of time series of fires, the objective is to develop predictive models capable of anticipating the occurrence of these events, assisting in environmental management and decision-making. By identifying patterns, trends and the main factors that influence fires, the research contributes to preventing and fighting fires, mitigating the impacts of climate change and promoting sustainability.

Keywords: Brazil, fires, data analysis, predictive models, sustainability.

1. Introdução

No contexto brasileiro os incêndios florestais representam um grave problema ambiental com impactos socioeconômicos e ambientais significativos. O Brasil, com sua vasta biodiversidade e riqueza em recursos naturais, enfrenta desafios significativos relacionados à gestão ambiental, em especial no que diz respeito aos incêndios florestais.

A ocorrência de queimadas, muitas vezes associada a fatores climáticos, mudanças no uso da terra e práticas agrícolas não sustentáveis, contribui para a degradação de ecossistemas, perda de biodiversidade, emissão de gases de efeito estufa e intensificação das mudanças climáticas. “Um estudo do Inpe mostra que as ondas de calor no Brasil saltaram de sete ocorrências para 32 nos últimos 30 anos.” (Redação National Geographic Brasil, 2024).

“Um estudo do Inpe (Instituto Nacional de Pesquisas Espaciais) que se debruçou em 60 anos de dados climáticos no Brasil mostrou que o número de ondas de calor no país aumentou mais de quatro vezes nos últimos 30 anos, tornando esses eventos climáticos extremos mais recorrentes e perigosos. Os dados também mostram que também em seis décadas a média de dias seguidos sem chuva saiu de 80 para 100, mostrando um cenário propício para temporadas de seca e refletindo a chegada da primavera 2024 no país como uma temporada seca e quente.” (Redação National Geographic Brasil, 2024).

“A instituição científica brasileira destaca que nas últimas três décadas, eventos climáticos como ondas de calor foram sete para 32, mostrando como o clima no Brasil pode estar mudando. A investigação foi feita a pedido do Ministério da Ciência, Tecnologia e Inovação (MCTI) com base em cálculos que mapearam o período de 1961 a 2020.” (Redação National Geographic Brasil, 2024).

Neste contexto, a análise de séries temporais de queimadas emerge como uma ferramenta relevante para compreender os padrões históricos, identificar tendências e prever futuros eventos. A presente pesquisa, enquadrada na área de Ciência de Dados e Análise Ambiental, tem como objetivo principal desenvolver um modelo preditivo para analisar séries temporais de queimadas no Brasil, utilizando dados de fonte pública do site do INPE dos eventos de 2019 a 2023, visando contribuir para a gestão ambiental e o desenvolvimento sustentável, possibilitando a implementação de medidas de prevenção e combate mais eficazes.

Este projeto se insere na intersecção da área de conhecimento entre a Ciência de Dados, especificamente a área de mineração de dados e aprendizado de máquina, e as Ciências Ambientais. A análise de séries temporais, uma técnica estatística fundamental para o estudo de dados coletados em intervalos de tempo regulares, é aplicada a um conjunto de dados de queimadas no Brasil, com o objetivo de extrair insights valiosos para a gestão ambiental contribuindo com uma sociedade mais segura.

O problema central deste estudo consiste em analisar as séries temporais de queimadas no Brasil, identificando padrões, tendências e fatores que influenciam a ocorrência desses eventos. Através da aplicação de técnicas de mineração de dados, busca-se desenvolver modelos preditivos capazes de antecipar a ocorrência de incêndios, auxiliando na tomada de decisões para a prevenção e o combate a esses eventos.

A escolha deste tema está diretamente relacionada ao Objetivo de Desenvolvimento Sustentável 13: “Ação contra a mudança global do clima”. As

queimadas contribuem significativamente para o aumento das emissões de gases de efeito estufa, intensificando o aquecimento global e suas consequências. Ao desenvolver modelos preditivos e identificar os fatores que influenciam a ocorrência de incêndios, este projeto contribui para o desenvolvimento de estratégias mais eficazes para mitigar os efeitos das mudanças climáticas e proteger o meio ambiente.

2. Motivações e Justificativas

A análise de séries temporais de queimadas no Brasil representa uma ferramenta poderosa para o avanço do conhecimento científico e para a solução de problemas ambientais. Ao fornecer informações precisas e oportunas sobre a ocorrência de incêndios, essa abordagem contribui para a gestão mais eficiente dos recursos naturais, a proteção da biodiversidade e a mitigação das mudanças climáticas.

Este estudo apresenta diversos interesses e benefícios, como o Conhecimento Científico, no qual permite aprofundar sobre os padrões de ocorrência de queimadas, a influência de fatores climáticos e socioeconômicos e a dinâmica dos ecossistemas em resposta ao fogo. É possível também obter o suporte à tomada de decisões, fornecendo informações cruciais para a elaboração de políticas públicas de prevenção e combate a incêndios, auxiliando na definição de áreas prioritárias para monitoramento e intervenção. Temos também o Gerenciamento de recursos naturais, onde se permite otimizar o uso de recursos para o combate a incêndios, direcionando esforços para as regiões mais vulneráveis e em períodos de maior risco. Podemos também abordar a Mitigação das mudanças climáticas, no qual as queimadas são uma importante fonte de emissões de gases de efeito estufa, contribuindo para o agravamento das mudanças climáticas. A análise de séries temporais pode auxiliar na quantificação dessas emissões e no desenvolvimento de estratégias para sua redução.

No quesito de avaliação do potencial de aplicabilidade da solução, entendemos que a solução proposta, baseada na análise de séries temporais, possui um alto potencial de aplicabilidade em diversas áreas, sendo elas: Gestão ambiental, Agricultura, Proteção civil, Pesquisa científica, Avaliação de políticas públicas e o Monitoramento em tempo real. A análise de séries temporais pode ser combinada com outras técnicas, como o geoprocessamento e o aprendizado de máquina, para gerar resultados ainda mais robustos e completos.

3. Objetivo

O projeto de Análise de Séries Temporais de Queimadas no Brasil tem como objetivo central aprofundar a compreensão dos padrões e tendências das queimadas no território nacional, a fim de desenvolver ferramentas para a gestão ambiental e a promoção da sustentabilidade. As principais pretensões e metas a serem alcançadas com

este estudo, utilizando métodos de aprendizado de máquina, análise exploratória, dentre outras metodologias da Ciências de Dados são:

- ✓ Aumentar o conhecimento científico sobre as causas e os impactos das queimadas no Brasil.
- ✓ Desenvolver ferramentas para auxiliar na gestão ambiental e na tomada de decisões.
- ✓ Contribuir para a construção de um futuro mais sustentável para o país.

Será realizado uma compreensão aprofundada do fenômeno através da identificação de padrões temporais e espaciais, mapeando a distribuição temporal e espacial das queimadas, identificando períodos de maior ocorrência, regiões mais afetadas e a relação com fatores climáticos e socioeconômicos. Análise de tendências, avaliando se há um aumento ou diminuição na frequência e intensidade das queimadas ao longo do tempo, e identificar possíveis causas para essas variações. Relação com variáveis climáticas e socioeconômicas para estabelecer a relação entre as queimadas e variáveis como temperatura, precipitação, índice de vegetação, uso do solo e atividades humanas.

O desenvolvimento de modelos preditivos será primordial na construção de modelos robustos, utilizando técnicas de mineração de dados para construir modelos preditivos capazes de antecipar a ocorrência de queimadas com alta precisão, utilizando a linguagem de programação python e o recurso colab. A incorporação de múltiplas variáveis, para aumentar a capacidade preditiva dos modelos. E também a avaliação da performance dos modelos, utilizando métricas de avaliação adequadas para medir a precisão e a confiabilidade dos modelos desenvolvidos.

4. Descrição da Base de Dados

Para o desenvolvimento deste projeto de Análise de Séries Temporais de Queimadas no Brasil, foi escolhido um dataset público no site da INPE (Instituto Nacional de Pesquisas Espaciais), <<https://terrabrasilis.dpi.inpe.br/queimadas/bdqueimadas/#exportar-dados>> considerando as queimadas que ocorreram entre os anos de 2019 a 2023. A base de dados possui 12 atributos, onde listamos cada um e uma breve descrição a que se refere dentro do contexto de estudo.

- Data / Hora – Registro da data e hora do evento.
- Satélite – Descrição do satélite onde ocorreu a queimada, “cujos dados diários de focos detectados são usados para compor a série temporal ao longo dos anos e assim permitir a análise de tendências nos números de focos para mesmas regiões e entre regiões em períodos de interesse.” (INPE, 2024).
- País – Nome do país do evento.
- Estado – Nome do estado do evento.
- Município – Nome do município do evento.
- Bioma – “É um conjunto de vida vegetal e animal, constituído pelo agrupamento de tipos de vegetação que são próximos e que podem ser identificados em nível regional, com condições de geologia e clima semelhantes e que, historicamente,

sofreram os mesmos processos de formação da paisagem, resultando em uma diversidade de flora e fauna própria”. (IBGE Educa, 2024).

- N. Dias Sem Chuva – Informa quantidade de dias sem chuva.
- Precipitação – Descreve o tipo de fenômeno relacionado à queda de água do céu. Isso inclui neve, chuva e chuva de granizo.
- Risco Fogo – Descreve o nível de Risco de Fogo da respectiva Região. “O princípio da estimativa meteorológica do Risco de Fogo (RF) para queima da vegetação é o de que quanto mais dias seguidos sem chuva em um local, maior seu risco de queima. Adicionalmente, são considerados no cálculo os efeitos locais do tipo da vegetação e do ciclo natural de seu desfolhamento, da temperatura máxima e umidade relativa mínima do ar diárias, da elevação topográfica e latitude, assim como a presença de fogo na área de interesse.” (Setzer, Sismanoglu, Santos, 2024).
- Latitude – É uma coordenada que especifica a posição norte-sul de um ponto na superfície da Terra ou de outro corpo celeste. A latitude é dada como um ângulo que varia de -90° no polo sul a 90° no polo norte, com 0° no Equador.
- Longitude – É uma coordenada geográfica que especifica a posição leste-oeste de um ponto na superfície da Terra, ou de outro corpo celestial. É uma medição angular, geralmente expressa em graus e denotada pela letra grega lambda.
- FRP – “Fire Radiative Power ou Potência Radiativa do Fogo, mede a energia radiante liberada por unidade de tempo, associada à taxa de queima da vegetação, em MW - megawatts. Esse valor é derivado a partir da comparação da energia medida no canal termal do pixel com fogo no seu interior com os pixels de seu entorno”. (INPE, 2024).

Esta base de dados não possui dados sensíveis e está de acordo com a Lei Geral de Proteção de Dados Pessoais – LGPD.

5. Referencial Teórico

Existem trabalhos que foram fundamentais para o desenvolvimento deste projeto, como o estudo de Sharma et al. (2024), que aborda modelos estatísticos e de aprendizado de máquina para prever incêndios e outros eventos de emergência na cidade de Edmonton.

A relevância deste trabalho para o presente estudo se justifica pela similaridade temática, que inclui a análise de séries temporais e o desenvolvimento de um modelo de aprendizado de máquina, além do período analisado, que contempla o impacto da pandemia de COVID-19 sobre as queimadas na região. O modelo desenvolvido no estudo foi baseado na análise de regressão, em que a heterogeneidade de Poisson é modelada utilizando a distribuição Gama. Esta distribuição é um modelo de dois parâmetros (média e dispersão), e suas principais vantagens incluem:

- ✓ Flexibilidade, permitindo um melhor ajuste quando a variância dos dados não é igual à média;
- ✓ Fácil interpretação, dado que a Gama possibilita um diagnóstico direto da sobre dispersão dos dados;
- ✓ Capacidade de modelar sobre dispersão, situação em que a variância é maior que a média.

No entanto, as principais limitações envolvem a complexidade computacional, maior do que a de um modelo Poisson simples, e a dificuldade em lidar com sub dispersão, quando a média é maior que a variância. Além disso, a suposição de que a heterogeneidade segue uma distribuição Gama pode não se adequar a todos os conjuntos de dados, dificultando a captura correta da estrutura dos dados.

Outro estudo importante para o desenvolvimento deste projeto foi o de Sun et al. (2024), que desenvolveu um modelo de predição de incêndios florestais baseado em autômatos celulares e aprendizado de máquina. A relevância deste estudo se dá pelo tema correlato, envolvendo a previsão de queimadas em florestas, com a utilização de séries temporais e múltiplas variáveis para a construção de modelos preditivos baseados em aprendizado de máquina. O estudo se baseou no modelo Wang Zhengfei, desenvolvido a partir de experimentos de ignição e princípios físicos relacionados à combustão florestal, incorporando o uso de autômatos celulares para prever a direção e a velocidade da propagação do fogo.

A abordagem de aprendizado de máquina adotada no estudo foi uma combinação de vários modelos, incluindo Support Vector Machine (SVM), Regressões Ridge e Lasso, além de incorporar testes de multicolinearidade para otimizar o modelo. Essa abordagem permite lidar com problemas complexos relacionados às queimadas florestais e apresenta como vantagens a capacidade de resolver tanto problemas simples quanto complexos, explorando diferentes aspectos dos dados. No entanto, as limitações incluem a complexidade do modelo, o que dificulta sua interpretação, e a alta demanda computacional, resultando em custos financeiros elevados para sua implementação.

Principais Conceitos dos Dados Analisados

1. Satélites são dispositivos lançados em órbita ao redor da Terra para monitoramento remoto e coleta de dados. No contexto da análise de queimadas, eles são essenciais para detectar focos de incêndio e medir temperaturas climáticas e ambientais. Satélites como os da série NOAA e AQUA/TERRA (MODIS) são amplamente utilizados para monitoramento ambiental e geração de imagens de focos de calor (LU, 2018).

2. Biomas são grandes ecossistemas que reúnem clima, fauna e flora semelhantes. No Brasil, os principais biomas afetados por queimadas são a Amazônia, o Cerrado e a Caatinga, cada um com características específicas de vulnerabilidade e respostas ao fogo (IBGE, 2004). Uma análise de queimadas por bioma permite compreender como diferentes ecossistemas respondem a incêndios e práticas de uso do solo.

3. Precipitação refere-se à quantidade de chuva que cai em uma determinada área e período, afetando diretamente a ocorrência de queimadas. Períodos prolongados de seca aumentam o risco de fogo em áreas florestais (ALVARES et al., 2013). A análise da relação entre a precipitação e os incêndios florestais é essencial para prever e entender a sazonalidade desses eventos.

4. Risco de fogo é um índice que indica a probabilidade de ocorrência de queimadas com base em variáveis como umidade do ar, temperatura, vento e a quantidade de material combustível presente (ALENCAR et al., 2015). Este indicador é fundamental para o planejamento de estratégias de prevenção de incêndios e alocação de recursos.

5. Latitude e longitude são regiões geográficas que definem a localização exata de um ponto na superfície da Terra. No contexto das queimadas, esses dados são essenciais para mapear os focos de incêndio e estudar sua distribuição espacial em diferentes regiões do Brasil (SNYDER, 2019).

6. Fire Radiative Power (FRP) é uma medida da energia radiativa liberada por um incêndio. Ela foi obtida a partir de sensores de satélite e correlacionada com a intensidade do fogo e a quantidade de biomassa queimada. O FRP é amplamente utilizado em estudos de queimadas para quantificar a gravidade dos incêndios (KAUFFMAN et al., 2003).

6. Pipeline da Solução

A análise de séries temporais de queimadas é importante para entender padrões, identificar tendências, impactos ambientais e tomar decisões mais informadas na gestão ambiental.

Esta construção requer uma construção bem estruturada para que seja uma entrega eficaz na solução que está sendo proposta, no entanto apresentamos um pipeline com as respectivas etapas que serão percorridas, particionando cada uma delas conforme descrevemos a seguir e como está representado na Figura 1.

1. Definição do Problema e Objetivos:

Problema: Compreender os padrões temporais e espaciais das queimadas no Brasil, identificar os principais fatores influenciadores e prever futuras ocorrências.

Objetivos: Construir um modelo preditivo para estimar a quantidade de focos de queimada em diferentes regiões do Brasil. Identificar os principais fatores climáticos, socioeconômicos e ambientais associados à ocorrência de queimadas. Analisar a evolução das queimadas ao longo do tempo e suas implicações para o meio ambiente e a sociedade.

2. Coleta e Preparação dos Dados:

Fontes de dados: INPE: Dados de focos de queimada.

Preparação dos dados: Limpeza, tratamento de dados faltantes, outliers e inconsistências. Transformação: Criação de novas variáveis (índices climáticos, variáveis socioeconômicas normalizadas). Agregação: Agregação dos dados por região e período de tempo.

3. Análise Exploratória de Dados (EDA):

Visualização: Gráficos de linhas, barras, mapas para visualizar a distribuição temporal e espacial das queimadas. Estatística descritiva: Cálculo de médias, desvios padrão, correlações para entender as características dos dados. Análise de componentes principais (PCA): Redução da dimensionalidade para identificar os principais padrões de variabilidade.

4. Modelagem:

Escolha do modelo de Séries temporais: ARIMA, SARIMA, modelos de espaço de estados. Machine learning: Random Forest, XGBoost, redes neurais para modelos mais complexos. Engenharia de features: Criação de novas features a partir dos dados originais para melhorar o desempenho do modelo. Validação: Divisão dos dados em treinamento, validação e teste, utilizando métricas como RMSE, MAE e accuracy.

5. Interpretação dos Resultados:

Análise dos coeficientes do modelo: Identificar os principais fatores que influenciam a ocorrência de queimadas. Análise de sensibilidade: Avaliar a robustez do modelo em relação a diferentes cenários. A criação de mapas de risco: Identificar as regiões mais propensas a ocorrências de queimadas.

6. Resultados:

Relatório: Apresentação clara e concisa dos resultados, incluindo gráficos, tabelas e mapas. Apresentação: Exposição dos principais achados e suas implicações para a gestão ambiental.

Figura 1: Pipeline



Elaborado pelo autor, 2024.

7. Análise Exploratória

As principais etapas que compõe a nossa análise exploratória de dados são:

1. Dados

Nesta fase inicial, o foco está na importação, tratamento e organização dos dados para uma análise mais consistente e estruturada. As etapas realizadas foram:

Importação de Dados: O conjunto de dados foi importado para um DataFrame a fim de possibilitar o tratamento e manipulação necessários.

Identificação de Dados para Tratamento: Nesta etapa, realizamos uma análise preliminar para identificar dados incompletos, inconsistentes ou que sejam reveladas de tratamento adicional, como valores nulos ou duplicados.

Visualização Inicial dos Dados: Foi gerado um gráfico inicial que possibilitou uma visualização prévia das variáveis, ajudando a verificar a presença de outliers e a distribuição geral dos dados.

Tratamento de Dados: As ações tomadas nesta etapa incluíram:

- ✓ Conversão de colunas com formato de dados para o tipo datetime.
- ✓ Criação de novas colunas baseadas em dados, quando necessário.
- ✓ Remoção de colunas constantes, que não apresentam variabilidade.
- ✓ Exclusão de valores específicos considerados como outliers, como -999 na coluna de probabilidade de fogo e dias sem chuva.
- ✓ Exclusão de valores nulos.
- ✓ Extração de informações adicionais de dados, como dia, semana e mês, para facilitar análises futuras.

Identificação e Tratamento de Outliers: Para garantir a qualidade dos dados, foi utilizada a técnica DBSCAN para identificar e remover outliers que pudessem distorcer as análises. A porcentagem de dados afetados por essa remoção foi documentada.

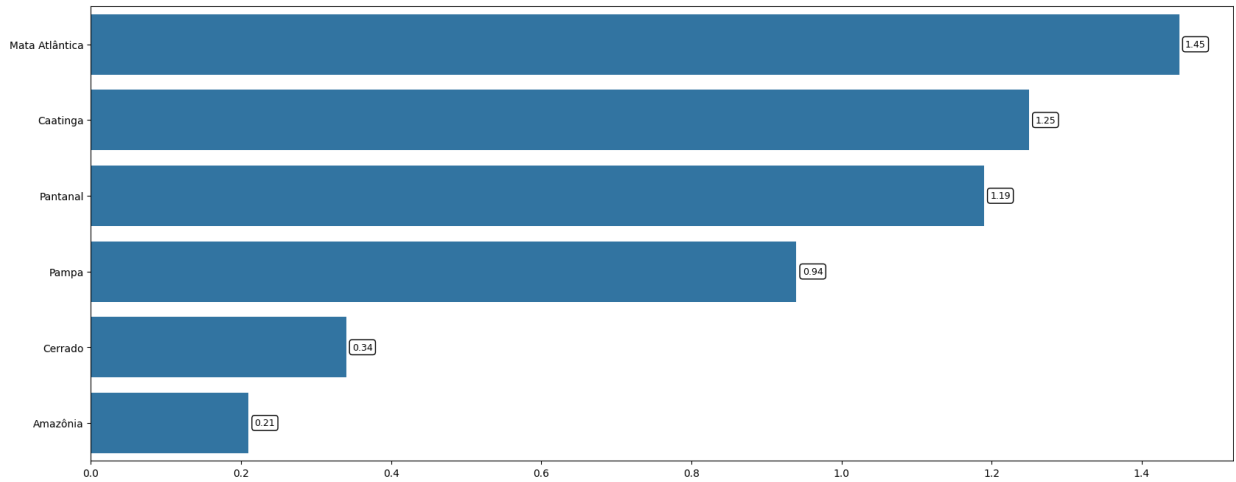
A análise apresenta-se em:

- **Outliers pontuais:** Podemos ver que nos dias sem chuva no Município Acre e com Bioma Amazônia temos dados que se destacam individualmente e ocorrem em um ponto específico sem relação com outros fatores, ou seja dias de sem chuva que esta fora do padrão em um período.
- **Outliers de contexto:** No seu caso, um dia com alta precipitação durante uma época de seca seria um outlier contextual. Um dia com alta precipitação durante uma época de seca seria um outlier contextual.

O modelo DBSCAN utilizado para remover outliers consegue detectar esses dois tipos de outliers na nossa base através da distancia dos registros na matriz de dados.

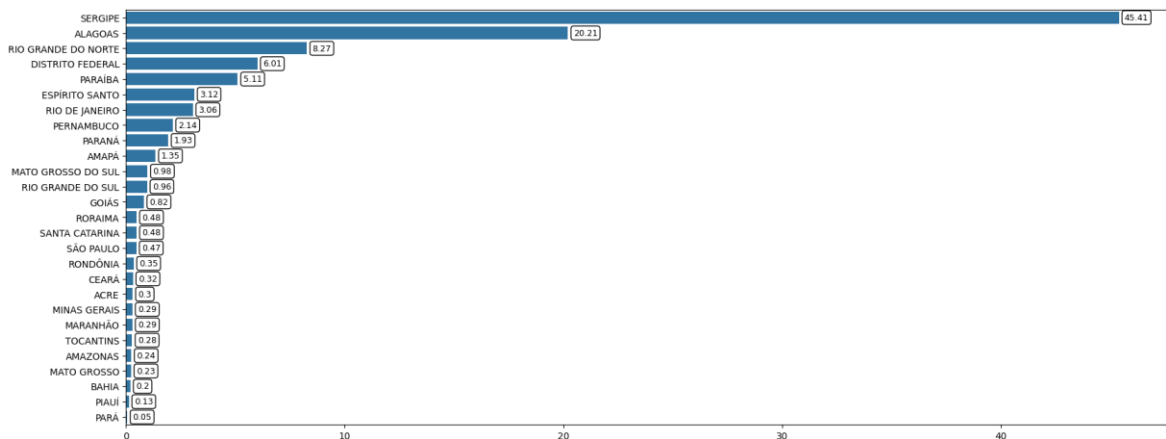
Agrupamento dos Dados (Group By): Para análises específicas, os dados foram agrupados por diferentes categorias, como bioma, representado na Figura 2 e estado, representado na Figura 3, facilitando a identificação de padrões por região.

Figura 2: Porcentagem Bioma



Elaborado pelo autor, recurso Colab, 2024.

Figura 3: Porcentagem Bioma



Elaborado pelo autor, recurso Colab, 2024.

2. Plots

Os gráficos utilizados nesta análise a fim de compreender e melhorar a qualidade dos dados foram:

Médias móveis com janela de 90 dias, sendo possível analisar as variações e tendências de número de dias sem chuvas, média de precipitação e média de risco de incêndios.

A janela de 90 dias permite captar variações sazonais e oferecer uma visão mais estável das condições climáticas, tornando possível uma análise mais precisa das tendências em períodos mais extensos.

- **Dias sem chuva:** verificar tendências de períodos prolongados e entender a frequência ao longo do tempo.
- **Precipitação:** verificar a variabilidade na quantidade de chuva e observar padrões sazonais.
- **Risco de fogo:** verificar períodos com alta ou baixa probabilidade de incêndios.

Correlação de Spearman, que torna possível analisar como as variáveis estão relacionadas entre si. Após essa análise foi notória a baixa correlação entre a maioria dessas variáveis, porém, algumas como o “RiscoFogo” e “DiaSemChuva” tiveram uma alta correlação positiva.

Spearman utiliza rankings dos dados, que permite criar relações de dados que não seguem uma distribuição normal e mais adequada para capturar relações não lineares.

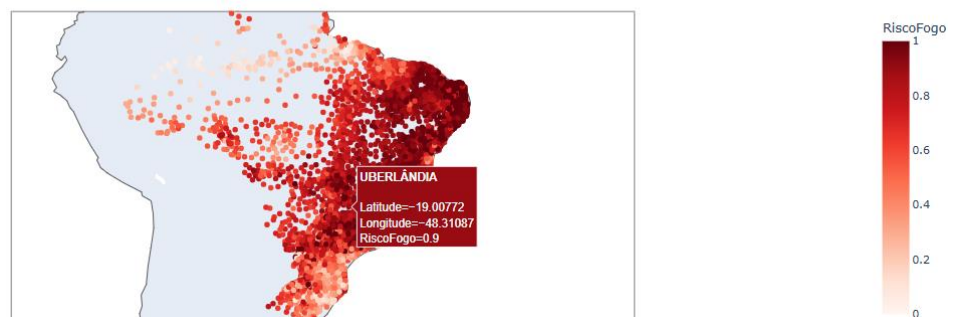
As variáveis risco de fogo e dias sem chuva possuem uma correlação de 0.7 o que indica uma correlação forte e positiva. Ou seja, conforme o número de dias sem chuva aumenta, o risco de fogo também tende a aumentar de forma significativa. Esse resultado é esperado, pois períodos secos aumentam a vulnerabilidade à ocorrência de incêndios, já que a baixa umidade e a falta de precipitação deixam o ambiente mais suscetível a queimadas.

No gráfico de mapa de calor, representadas nas Figuras 4 e 5, podemos observar que as regiões Sudeste e Nordeste do Brasil apresentam um maior risco de fogo em comparação com o restante do país. Essas áreas também registram mais dias sem chuva, o que contribui para a elevação do risco de incêndios.

Esse padrão indica uma vulnerabilidade climática nessas regiões, onde os períodos prolongados de seca e a baixa precipitação criam condições propícias para a ocorrência de incêndios. A visualização em mapa facilita a identificação espacial dessas tendências, permitindo um entendimento mais claro de como as condições climáticas variam entre diferentes áreas do Brasil e onde estão concentradas as maiores áreas de risco.

Figura 4: Média de Risco Fogo

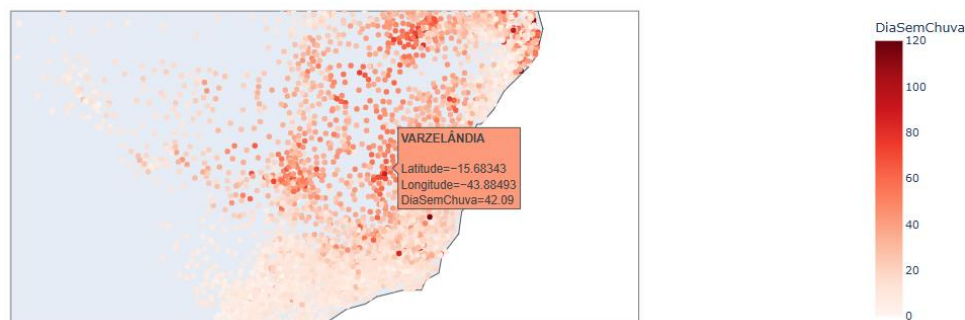
RiscoFogo - Média



Elaborado pelo autor, recurso Colab, 2024.

Figura 5: Média dias sem chuva

DiaSemChuva - Média



Elaborado pelo autor, recurso Colab, 2024.

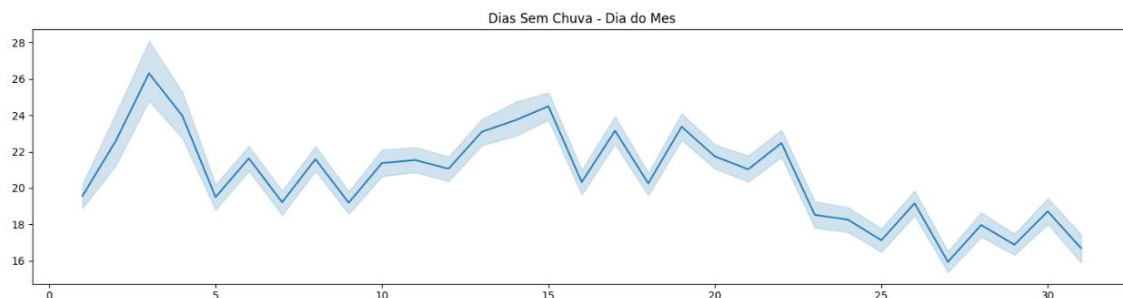
Decomposição de séries temporais, de forma a verificar a tendência, sazonalidade e ruído. Foram verificadas a decomposição sazonal por 180 e por 30 dias, além da decomposição para diferentes combinações entre estado e bioma presentes na data frame, a média de dias sem chuva foi utilizada como a variável de interesse;

Dispersão e BoxPlot, conforme representado nas Figuras 6 e 7, de forma a verificar o comportamento, distribuição e variação dos dias sem chuvas em diferentes períodos de tempo: dia, semana, mês e ano;

Autocorrelação e Autocorrelação parcial, a primeira mede a correlação temporal consigo mesma e a segunda entre a série temporal e suas defasagens, de forma a isolar o efeito direto de cada uma. Para estes gráficos foi usada a variável “media_risco_fogo”;

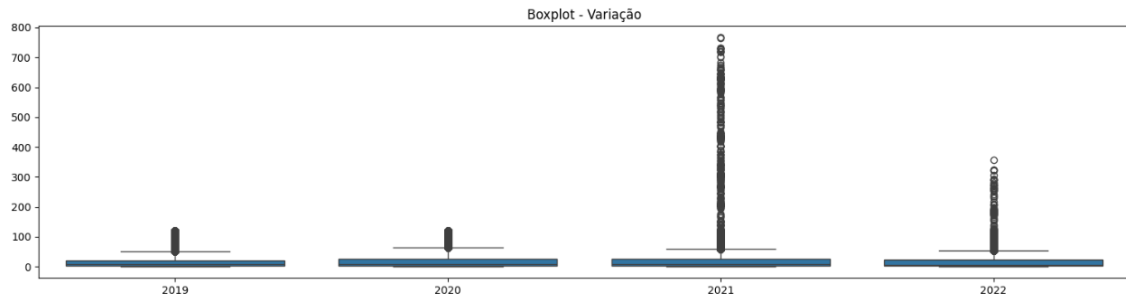
Teste de Dickey-Fuller, usado para medir a estacionaridade das séries temporais do risco de incêndios em regiões e biomas. Esta métrica é importante para verificar a previsibilidade do modelo e é fundamental para que o modelo ARIMA possa prever corretamente.

Figura 6: Dias do mês sem chuva



Elaborado pelo autor, recurso Colab, 2024.

Figura 7: Dias sem chuva por ano



Elaborado pelo autor, recurso Colab, 2024.

O Teste de Dickey-Fuller é amplamente utilizado em diversas áreas. A identificação de séries estacionárias é crucial para a modelagem de previsões, pois muitos modelos econométricos, como ARIMA (AutoRegressive Integrated Moving Average), assumem que os dados são estacionários. Além disso, o teste é frequentemente utilizado em estudos de Co integração, onde se busca entender a relação de longo prazo entre duas ou mais séries temporais.

O Teste de Dickey-Fuller é uma técnica essencial na análise de séries temporais, permitindo que pesquisadores e analistas determinem a estacionaridade de seus dados. Compreender a natureza dos dados é crucial para a modelagem e a previsão, e o teste fornece uma base sólida para essa análise. A aplicação correta do teste, juntamente com uma interpretação cuidadosa dos resultados, pode levar a insights valiosos e decisões informadas em uma variedade de campos.

Quando realizamos o teste de Dickey-Fuller na variável risco de fogo, o comportamento dos dados com valores entre 0 e 1 sugere que o teste pode identificar que a série temporal é estacionária se não houver tendência ou sazonalidade nos valores ao longo do tempo.

8. Metodologia

O modelo de previsão de risco de incêndio escolhido foi o ARIMA. Esta técnica ARIMA (Auto-Regressive Integrated Moving Average), ou Modelo Autorregressivo Integrado de Médias Móveis, é uma metodologia amplamente utilizada para previsão em séries temporais, especialmente em contextos de variações climáticas e de risco, como a análise de risco de incêndio. Esse modelo é composto por três parâmetros: o termo autorregressivo (AR), que considera os valores passados da série; o termo de integração (I), que transforma a série em estacionária por meio de diferenciações; e o termo de média móvel (MA), que utiliza o erro das variações anteriores para ajustar

Antes de ajustar o modelo ARIMA, é necessário verificar se a série temporal é estacionária, ou seja, se suas características estatísticas, como média e variância, não variam ao longo do tempo. Para esta análise, o Teste de Dickey-Fuller é frequentemente utilizado, pois permite identificar se há uma tendência significativa que comprometeria a previsibilidade do modelo (ANDRADE; ZAMBON, 2018). Quando uma série não é

estacionária, é aplicada uma técnica de diferenciação, que calcula a variação entre os dados consecutivos até que se obtenha uma série estável.

Os parâmetros do modelo ARIMA são definidas por (p, d, q), onde:

Seleção dos parâmetros

- **p** (autorregressão): Número de lags passados que influenciam o valor atual.
- **d** (diferenças): Números de vezes que a série precisa ser diferenciada para se tornar estacionária.
- **q** (médias móveis): Número de erros passados que influenciam o valor atual.

A escolha dos valores de p, d e q.

De acordo com Morettin e Toloi (2006), o modelo ARIMA é indicado para séries temporais com comportamento sazonal e tendências pronunciadas, características comuns em características naturais, como os riscos de incêndio associados a períodos de seca. Este modelo, ao capturar padrões históricos na série temporal de “RiscoFogo”, contribui para a identificação de variações e padrões que podem influenciar a segurança ambiental e o planejamento de ações preventivas em áreas de risco.

Utilizando a biblioteca estatística Python "pmdarima" muito popular para modelagem de séries temporais, pois oferece implementações eficientes e flexíveis de modelos ARIMA e SARIMA. Ele é especialmente útil para:

1. Seleção automática de modelos: O pmdarima pode auxiliar na escolha dos melhores hiper parâmetros para modelos ARIMA, economizando tempo e esforço.
2. Forecasting: Permite realizar previsões precisas de séries temporais, considerando a sazonalidade e tendências.
3. Análise de séries temporais: Oferece ferramentas para explorar e analisar as características de séries temporais.

Técnicas aplicadas na implementação do algoritmo: “AUTO-ARIMA W/ SEASONAL - WEEKLY (SARIMA)”

Auto_arima foi configurado para encontrar automaticamente a melhor combinação de parâmetros para um modelo ARIMA sazonal com variáveis exógenas.

exogenous: x_train

- Variáveis externas que influenciam a série temporal.

y: y_train

- Valores observados da série temporal a ser modelada.

start_p: 1

- Valor inicial da ordem da parte autoregressiva do modelo (AR).

start_q: 1

- Valor inicial da ordem da parte de média móvel do modelo (MA).

max_p: 3

- Valor máximo da ordem da parte autoregressiva do modelo (AR).

max_q: 3

- Valor máximo da ordem da parte de média móvel do modelo (MA).

start_P: 0

- Valor inicial da ordem autoregressiva sazonal.

seasonal: True

- Modelo será sazonal?

m: 52

- Número de períodos em uma temporada sazonal.

d: 1

- Número de vezes que o modelo irá diferenciar a série temporal uma vez para torná-la estacionária.

D: 1

- Número de vezes que o modelo realizará uma diferenciação sazonal para capturar as variações sazonais na série temporal.

trace: trace

- Monitorar ajustes do modelo.

error_action: ignore

- Como o modelo lida com erros durante o ajuste.

suppress_warnings: True

- Suprimir avisos durante o processo de ajuste do modelo?

stepwise: True

- Otimizar progressivamente os valores de p, q, P, Q?

9. Resultados e discussão

Os aspectos dos resultados obtidos antecedem a toda análise realizada no processo de construção do modelo, avaliando os pontos positivos identificados refere-se à qualidade dos dados, pois a base de dados foi processada com etapas de limpeza robustas, incluindo o tratamento de valores nulos e remoção de outliers com DBSCAN. A agregação por períodos (diária, semanal) e diferentes níveis geográficos (Estado, Bioma) proporcionou granularidade na análise.

Gráficos de séries temporais para médias móveis e análises sazonais revelaram padrões de longo prazo, além de identificar picos anômalos de risco de fogo.

Boxplots de variáveis como "Dias Sem Chuva" e "Precipitação" permitiram entender as variações ao longo de períodos (dias, semanas, meses e anos).

A análise de correlação de Spearman foi útil para identificar relações entre variáveis climáticas e o "Risco de Fogo".

O uso do teste de Dickey-Fuller foi apropriado para avaliar a estacionaridade das séries temporais, requisito fundamental para a aplicação do modelo ARIMA.

A seleção automática do melhor modelo ARIMA (Auto-Arima) garantiu eficiência ao balancear complexidade e desempenho.

A previsão semanal demonstrou capacidade de capturar tendências de risco de incêndio em curto e médio prazo.

Os gráficos gerados (ACF, PACF, decomposição da série) facilitaram a interpretação de padrões sazonais e autocorrelações, fornecendo insights valiosos para políticas públicas.

Já nos aspectos negativos e as limitações observadas são os fatores externos não considerados, como: Variáveis adicionais, como densidade populacional, uso da terra e atividades humanas (desmatamento, queimadas intencionais), não foram incluídas, o que pode limitar a explicação dos resultados. A necessidade de ajustes para tornar as séries estacionárias pode indicar que alguns padrões importantes (como sazonalidade) não foram completamente capturados.

O ARIMA é eficaz para séries estacionárias e uni variadas, mas não considera relações entre múltiplas variáveis (como "Precipitação" e "Dias Sem Chuva"). Modelos mais avançados, como SARIMA ou LSTM (redes neurais), poderiam ser explorados.

A divisão por biomas e estados é interessante, mas as previsões podem sofrer com a falta de dados históricos robustos para algumas regiões, comprometendo a confiabilidade dos modelos.

O período analisado (2019–2023) pode ser curto para capturar padrões de longo prazo, especialmente aqueles influenciados por ciclos climáticos (como El Niño e La Niña). Erros ou lacunas na base de dados do INPE podem impactar os resultados, e a dependência de uma única fonte de dados pode ser uma fragilidade.

1. Entendimento do resultado do modelo de previsão:

AIC (Critério de Informação de Akaike), é uma métrica usada para avaliar a qualidade de modelos estatísticos, considerando a complexidade (número de parâmetros) e o ajuste aos dados.

Quanto menor o AIC, melhor o modelo, pois ele consegue explicar os dados de forma eficiente, minimizando o erro e a sobrecarga de parâmetros desnecessários.

O “BEST MODEL” indica a ordem do modelo ARIMA selecionado pelo algoritmo. A notação ARIMA (p, d, q) (P, D, Q) [m] representa:

p: Ordem do processo autorregressivo (AR).

d: Número de diferenças para tornar a série estacionária.

q: Ordem do processo de médias móveis (MA).

P, D, Q: Equivalentes a p, d e q, mas para a componente sazonal.

m: Período da sazonalidade (considerando, 52 semanas).

A estrutura do modelo ARIMA é definida a partir de:

- ✓ Parâmetros de autorregressão (AR) e média móvel (MA) para padrões locais.
- ✓ Ordem da diferenciação para tornar a série estacionária.
- ✓ Parâmetros sazonais para autorregressão e média móvel.
- ✓ Ordem de diferenciação sazonal.
- ✓ Período sazonal (aqui, 52 semanas para dados semanais).

O termo intercept indica a presença de um valor constante no modelo, ajustado conforme os dados.

2. Interpretação do resultado por Estado e Bioma:

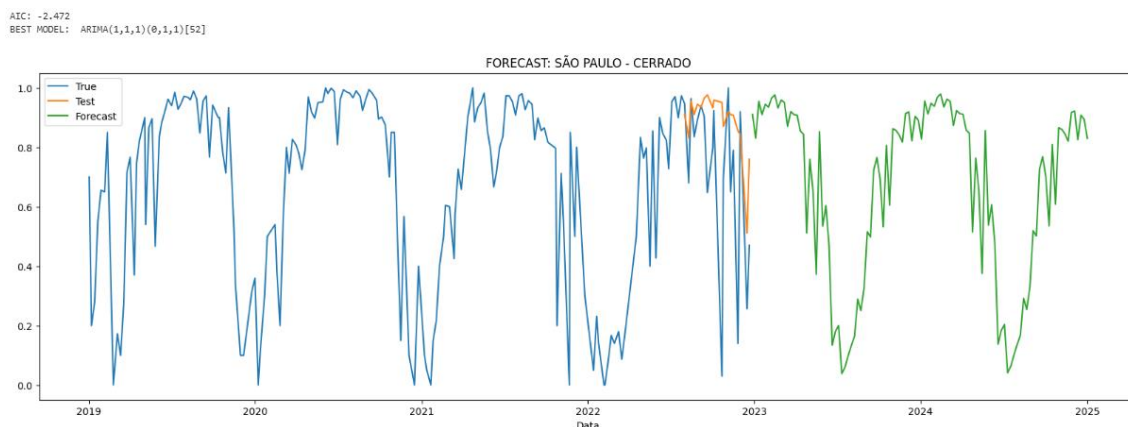
- **Dependência:** O valor $AR=1$ sugere que o número de incêndios em um determinado período depende do número de incêndios no período anterior.
- **Tendência:** A diferenciação ($I=1$) indica que há uma tendência crescente ou decrescente no número de incêndios ao longo do tempo.
- **Sazonalidade:** A componente sazonal $(0,1,1)[52]$ confirma a existência de um padrão sazonal anual, ou seja, o número de incêndios varia de acordo com as estações do ano.
- **Erros de previsão:** O termo $MA=1$ indica que os erros de previsão do modelo estão correlacionados com os erros de previsão anteriores.
- **Qualidade do ajuste:** O valor de AIC sugere que o modelo se ajusta razoavelmente bem aos dados, mas pode haver espaço para melhorias, especialmente se houver outros modelos com valores de AIC ainda menores.

Os casos com menor AIC representam modelos mais eficientes na sua previsão.

➤ São Paulo - Cerrado (AIC: -2.472):

O modelo, representado na Figura 8, $ARIMA(1,1,1) (0,1,1) [52]$ com AIC -2.472 um valor de AIC negativo indica que o modelo explica bem os dados. Quanto menor o AIC, melhor o ajuste do modelo aos dados. Sugere que o número de incêndios é influenciado por uma combinação de fatores, incluindo a ocorrência de incêndios em períodos anteriores, uma tendência de longo prazo e padrões sazonais anuais.

Figura 8: Estado São Paulo, Bioma Cerrado

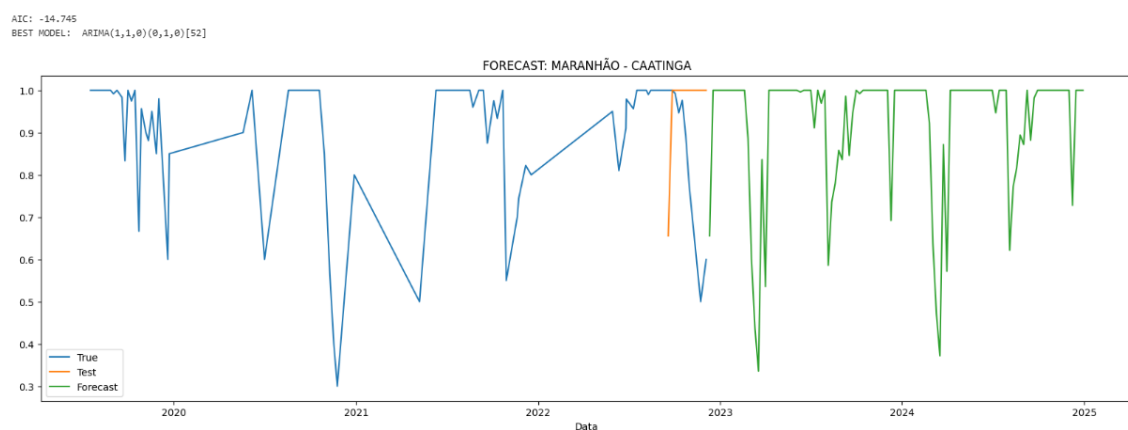


Elaborado pelo autor, recurso Colab, 2024.

➤ **Maranhão - Caatinga (AIC: -14.745):**

O modelo, representado na Figura 9, ARIMA(1,1,0)(0,1,0)[52] com AIC -14.745 sugere que o número de incêndios é influenciado por uma combinação de fatores, incluindo: Tendência: O número de incêndios está aumentando ou diminuindo ao longo do tempo. Sazonalidade: Há um padrão sazonal anual nos dados, com mais incêndios em determinadas épocas do ano. Autocorrelação: O número de incêndios em um determinado período está relacionado ao número de incêndios no período anterior.

Figura 9: Estado Maranhão, Bioma Caatinga



Elaborado pelo autor, recurso Colab, 2024.

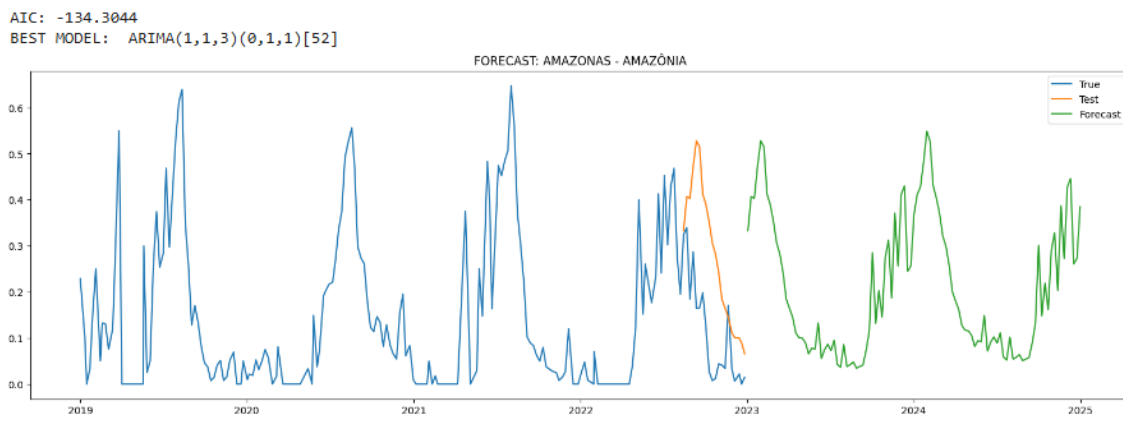
➤ **Amazonas - Amazônia (AIC: -134.3044):**

O modelo representado na Figura 10, indica que há uma relação significativa de dependência de curto prazo (ordem) e que os padrões sazonais (ordem sazonal) desempenham um papel relevante.

A alta eficiência sugere uma previsão confiável para esta região, onde o padrão sazonal pode estar associado ao regime de chuvas da Amazônia.

O modelo ARIMA (1,1,3) (0,1,1) [52] sugere que o número de incêndios é influenciado pelo valor anterior, pelos três últimos erros de previsão e pela sazonalidade anual, com um componente de média móvel sazonal.

Figura 10: Estado Amazonas, Bioma Amazônia

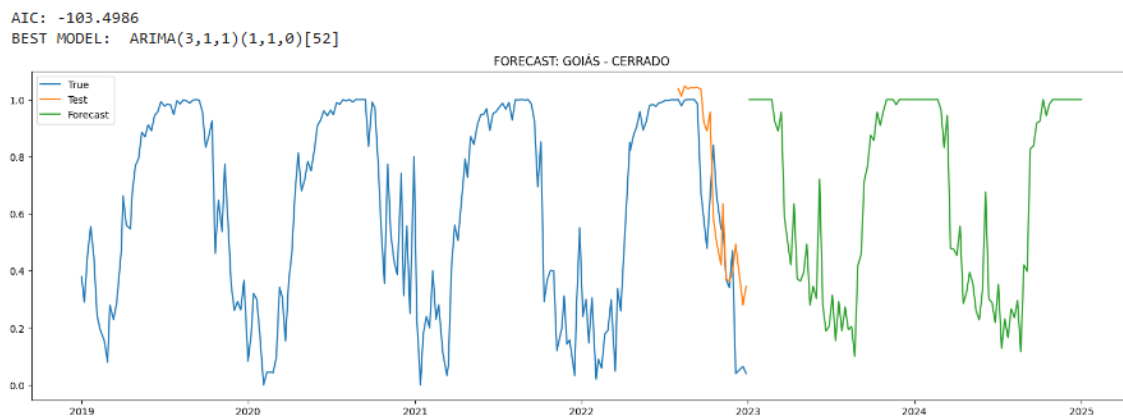


➤ Goiás - Cerrado (AIC: -103.4986):

Modelo captura dependências mais complexas com maior número de parâmetros. Conforme pode ser observado na Figura 11.

Isso pode refletir a variabilidade climática mais complexa do Cerrado, incluindo períodos secos prolongados.

Figura 11: Estado Goiás, Bioma Cerrado



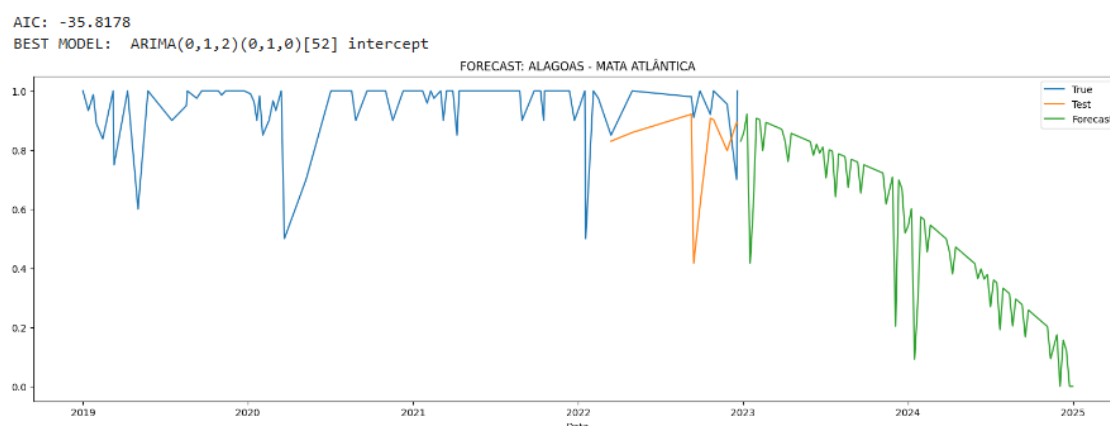
➤ **Alagoas - Mata Atlântica (AIC: -35.8178):**

Modelo simples, sem padrões sazonais significativos, representado na Figura 12.

Sugere que o risco de fogo na Mata Atlântica em Alagoas pode depender principalmente de eventos climáticos locais, sem padrões cíclicos claros.

O modelo ARIMA (0,1,2) (0,1,0) [52] indica que o número de incêndios é influenciado pelos dois últimos erros de previsão e pela sazonalidade anual.

Figura 12: Estado Alagoas, Bioma Mata Atlântica



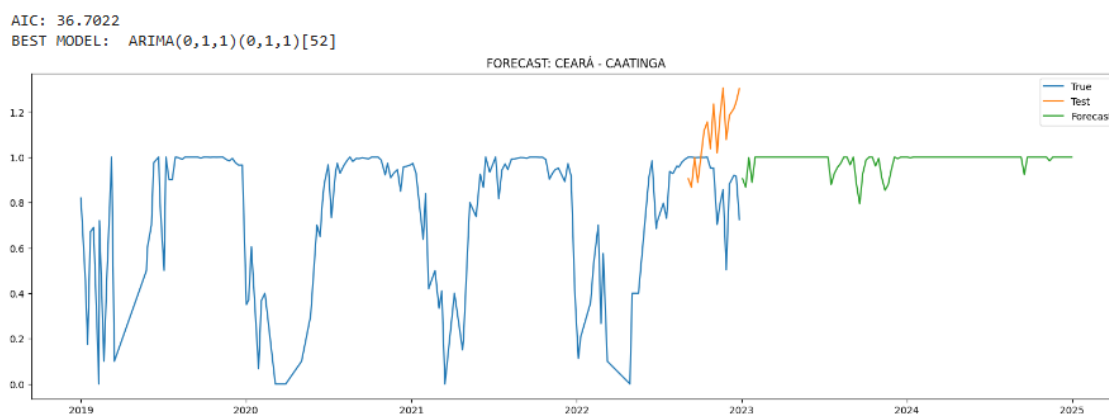
Elaborado pelo autor, recurso Colab, 2024.

➤ **Ceará - Caatinga (AIC: 36.7022):**

Modelo é relativamente simples, mas o AIC elevado indica que os dados são difíceis de modelar.

A alta variabilidade do regime hídrico da Caatinga, com chuvas concentradas em curtos períodos, pode dificultar a captura de padrões claros pelo modelo, como é observado na Figura 13.

Figura 13: Estado Ceará, Bioma Caatinga



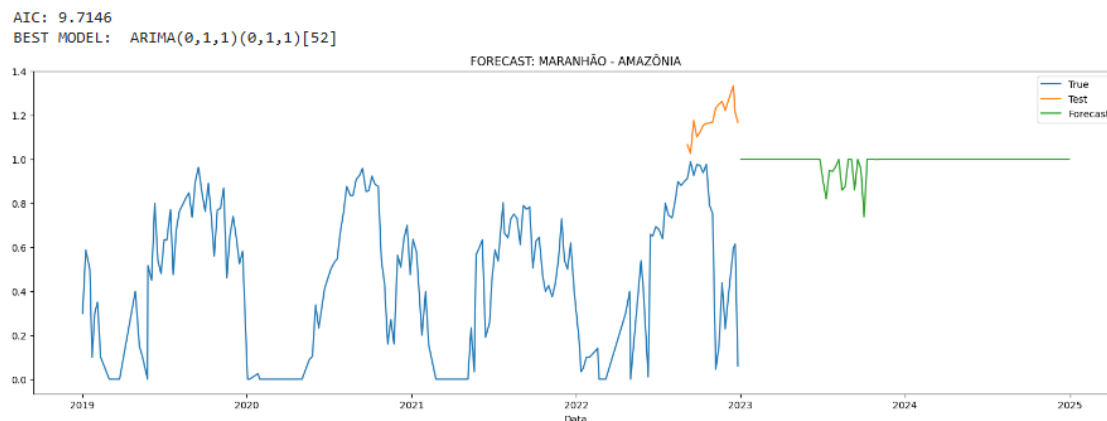
Elaborado pelo autor, recurso Colab, 2024.

➤ **Maranhão - Amazônia (AIC: 9.7146):**

O modelo sugere que os dados apresentam padrões sazonais, mas o AIC positivo indica que há espaço para melhorar o ajuste.

Isso pode ser atribuído à transição entre biomas no Maranhão (Amazônia e Cerrado), representado na Figura 14, que introduz maior complexidade.

Figura 14: Estado Maranhão, Bioma Amazônia



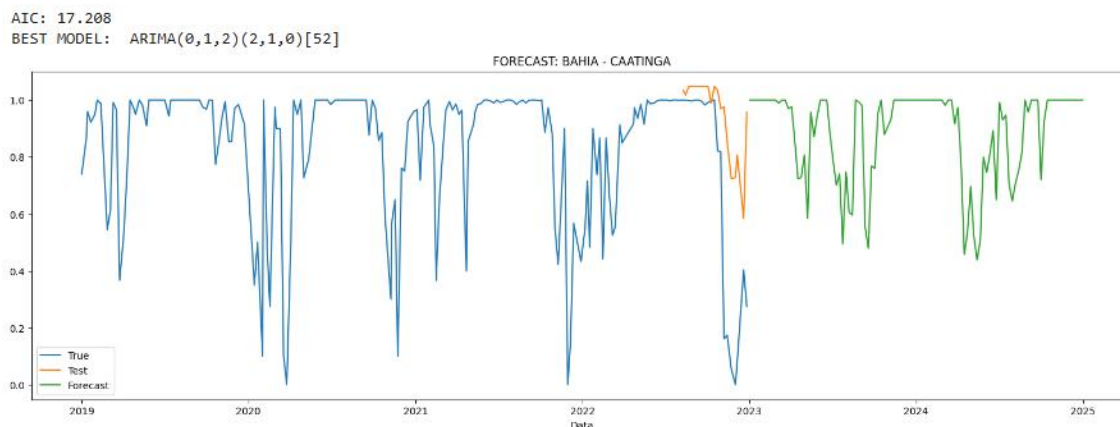
Elaborado pelo autor, recurso Colab, 2024.

➤ **Bahia - Caatinga (AIC: 17.208):**

O modelo representado na Figura 15, tenta capturar padrões de curto prazo e alguma sazonalidade.

O AIC positivo reflete dificuldades do modelo em lidar com as características específicas da Caatinga, como os períodos prolongados de seca.

Figura 15: Estado Bahia, Bioma Caatinga

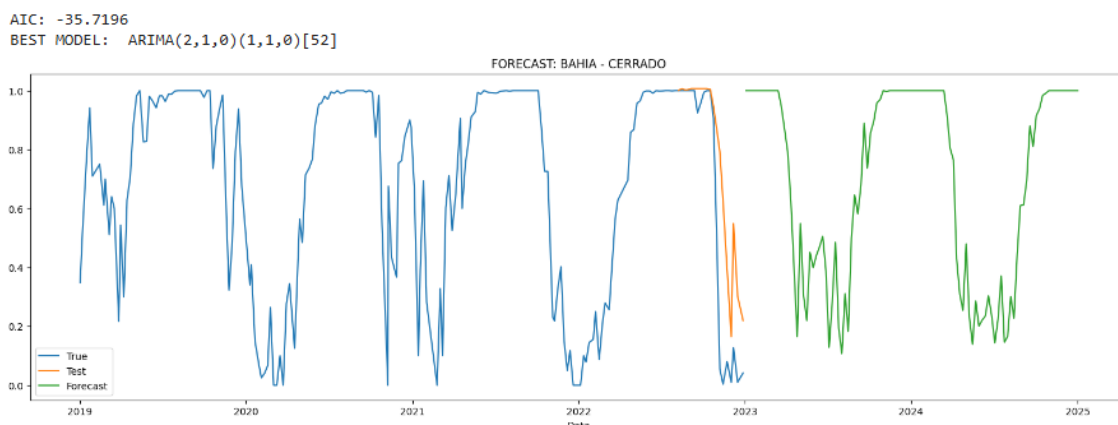


Elaborado pelo autor, recurso Colab, 2024.

➤ **Bahia - Cerrado (AIC: -35.7196) e Bahia - Mata Atlântica (AIC: -7.6786):**

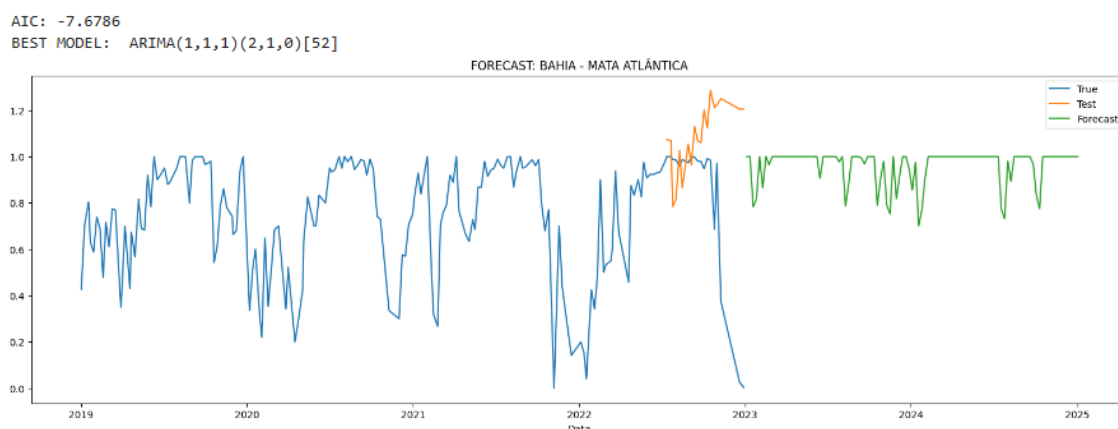
Estes modelos indicam desempenho moderado. A Bahia com bioma Cerrado, representado na Figura 16, e a Bahia com bioma Mata Atlântica, representado na Figura 17, apresentam diversidade climática e ecológica, o que pode justificar essa variação no ajuste.

Figura 16: Estado Bahia, Bioma Cerrado



Elaborado pelo autor, recurso Colab, 2024.

Figura 17: Estado Bahia, Bioma Mata Atlântica



Elaborado pelo autor, recurso Colab, 2024.

3. Análise conclusiva:

Modelos mais confiáveis apresentam modelos com AIC muito baixo, indicando que os padrões de risco de fogo nessas regiões são bem explicados pelos dados disponíveis. O modelo ajuda a entender as principais forças que influenciam a ocorrência de incêndios.

Dificuldades de modelagem, apresentam dificuldades maiores em capturar padrões claros, possivelmente devido à variabilidade climática ou à qualidade dos dados.

Dependência de sazonalidade, onde a sazonalidade desempenha um papel crucial, os modelos são mais simples e independem de ciclos sazonais claros. A presença de

termos sazonais (P, D, Q) em todos os modelos sugere uma forte componente sazonal nos dados, provavelmente relacionada a padrões climáticos anuais.

O valor de “d” igual a 1 em todos os modelos indica que a série temporal original precisou ser diferenciada uma vez para se tornar estacionária, sugerindo a presença de uma tendência nos dados.

Os modelos ARIMA variam significativamente entre os diferentes estados e biomas. Isso indica que as dinâmicas dos incêndios são influenciadas por fatores locais específicos, como clima, tipo de vegetação e práticas de manejo do fogo.

Limitações: AIC mais altos indicam que modelos ARIMA podem não ser suficientes para algumas regiões, sendo recomendável testar abordagens mais complexas, como SARIMA ou modelos baseados em aprendizado de máquina.

4. Avaliação de melhoria da aplicação técnica:

O que pode ser melhorado na precisão do modelo é testar modelos alternativos como o SARIMA que é capaz de capturar sazonalidade de forma mais precisa em séries temporais com ciclos sazonais complexos.

Modelos baseados em aprendizado de máquina, como: Redes neurais recorrentes (LSTM/GRU) para capturar dependências temporais não-lineares. Random Forest ou XGBoost para explorar interações não-lineares entre variáveis climáticas. Prophet, ideal para séries com sazonalidade forte.

A inclusão de novas variáveis para incorporar dados de temperatura, umidade relativa do ar e densidade populacional. Usar índices climáticos mais completos, como o NDVI (Índice de Vegetação por Diferença Normalizada) para captar variações na cobertura vegetal.

A validação cruzada também é uma sugestão para avaliar a robustez do modelo em diferentes períodos ou regiões.

Explorar modelos que considerem múltiplas variáveis simultaneamente, como VAR (Vetores Autorregressivos).

Realizar teste de diferentes combinações de para os modelos ARIMA/SARIMA usando técnicas de otimização automatizada.

5. Aplicação prática dos resultados do modelo:

✓ Na Gestão e Prevenção de Queimadas como:

Planejamento Regional: Direcionar esforços de prevenção para estados/biomas em períodos de maior risco.

Alocação de Recursos: Definir prioridades para o uso de brigadas, sistemas de monitoramento e campanhas educativas.

Alertas Precoces: Utilizar previsões de risco de fogo para emitir alertas em tempo hábil, ajudando na preparação de comunidades vulneráveis.

✓ **Políticas Públicas:**

Apoiar políticas públicas de mitigação de queimadas com base em evidências científicas.

Informar políticas de reflorestamento em áreas críticas.

Criar sistemas de alerta precoce, permitindo que as autoridades tomem medidas preventivas antes da ocorrência de incêndios.

As previsões podem ser utilizadas para comunicar os riscos de incêndio à população e sensibilizar para a importância da prevenção.

✓ **Empresas e Setores Privados:**

Empresas agrícolas podem ajustar suas atividades para mitigar riscos durante períodos críticos.

Concessionárias de energia podem usar os dados para prever riscos em linhas de transmissão em áreas de queimadas.

10. Conclusão

Este projeto aplicado visou compreender os padrões temporais das queimadas no Brasil, utilizando técnicas avançadas de análise de séries temporais e aprendizado de máquina. O problema de pesquisa abordou a identificação de fatores climáticos, socioeconômicos e ambientais associados às queimadas e a previsão da sua ocorrência, com o objetivo de subsidiar estratégias de gestão ambiental. Para isso, empregou-se um pipeline abrangente que incluiu desde a coleta e tratamento de dados até a modelagem preditiva com ARIMA.

Os resultados indicaram que o modelo ARIMA foi eficaz para captar padrões sazonais e tendências em algumas regiões, como Amazonas e Goiás, onde os modelos apresentaram baixo AIC, sugerindo alta precisão. No entanto, em outras áreas, como Ceará e Maranhão, houve dificuldades em capturar padrões claros devido à variabilidade climática ou limitações na qualidade dos dados. A análise demonstrou que fatores como precipitação, dias sem chuva e risco de fogo são variáveis-chave na previsão de queimadas, enquanto a sazonalidade desempenhou um papel crucial em regiões específicas.

A análise dos modelos ARIMA por estado e bioma permite identificar padrões e tendências nos dados de incêndios e pode ser utilizada para auxiliar na tomada de decisões relacionadas à gestão de riscos de incêndio. Ao combinar modelos estatísticos com conhecimento especializado, é possível desenvolver sistemas de previsão mais precisos e eficazes, contribuindo para a prevenção e o combate a incêndios florestais. No entanto, é importante lembrar que os modelos estatísticos são ferramentas de apoio e devem ser interpretados em conjunto com outras informações relevantes.

As contribuições deste estudo incluem a aplicação de métodos robustos de análise de séries temporais e a utilização de ferramentas como o Teste de Dickey-Fuller e a correlação de Spearman para melhorar a qualidade dos dados. Os insights obtidos podem auxiliar na formulação de políticas públicas mais eficazes no combate às queimadas. Contudo, limitações foram identificadas, como a ausência de variáveis socioeconômicas e de uso do solo, a dependência de dados univariados no modelo ARIMA e a necessidade de um maior período de análise para capturar padrões de longo prazo.

Este estudo representa um passo importante na modelagem preditiva de queimadas, destacando a importância de integrar abordagens metodológicas robustas e dados de alta qualidade para enfrentar os desafios ambientais do Brasil.

11. Referências bibliográficas

ALENCAR, A.; NEPSTAD, D.; CARVALHO, GO O papel das florestas tropicais na moderação do clima e no ciclo hidrológico. *Ciência e Cultura*, São Paulo, v. 4, pág. 56-64, 2015.

ALVARES, CA et al. Mapa de classificação climática de Köppen para o Brasil. *Meteorologische Zeitschrift*, v. 6, pág. 711-728, 2013.

IBGE. Mapa de biomas do Brasil: primeira aproximação. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística, 2004.

KAUFFMAN, JB et al. Fatores de consumo de biomassa e liberação de carbono para incêndios em florestas tropicais. *Ecology*, v. 84, n. 2, p. 351-370, 2003.

LU, D. Sensoriamento remoto do ambiente: uma perspectiva de recursos terrestres. 2. ed. Nova York: Pearson, 2018.

SNYDER, JP Projeções de mapas: um manual de trabalho. Washington: US Geological Survey, 2019.

ANDRADE, FM; ZAMBON, LM Análise de séries temporais: um modelo ARIMA aplicado ao setor energético brasileiro. *Revista Brasileira de Engenharia de Produção*, v. 2, pág. 170-188, 2018.

GOMES, CFS; CASSEL, RA Previsão em séries temporais: métodos tradicionais e aplicações no ambiente organizacional. S.

MORETTIN, P. A.; TOLOI, C. M. de C. Análise de séries temporais. 3. ed. São Paulo: Egard Blucher, 2018.

SOUSA, Alex R S.; SILVA, Cristiane; SILVA, Juliane SF; e outros. Análise de séries temporais. Porto Alegre: SAGAH, 2021.

12. Diretório GitHub

Todo o conteúdo do projeto estará disponível no site da GitHub, que poderá ser acessado pelo link:

<https://github.com/mucioferraz/Projeto-CLML>

O diretório está organizado por pastas:

Pasta “Análise Exploratoria” será disponibilizado os códigos em Python que foram utilizados para realizar a análise exploratória, tratamento dos dados e o estudo de séries temporais.

Pasta “Dados” temos os arquivos das bases de dados utilizados para o estudo.

Pasta “Documentos” temos o cronograma de entrega do projeto, as versões de entrega deste documento e as versões de entrega do colab:
https://github.com/mucioferraz/Projeto-CLML/blob/main/Projeto_aplicado_IV_GrupoCLML_entrega_4.ipynb

Temos também o arquivo README.md com informações relevantes do projeto.

Link do vídeo com a apresentação final do projeto e do modelo:
<https://www.youtube.com/watch?v=olW9IVjm5Js>