PROJETO APLICADO 4 - SÉRIES TEMPORAIS – modificado etapa4

**Resumo:**

Série Temporal é uma sequência de realizações e observações de uma variável ao longo do tempo e registrado em períodos regulares. É realizada com o intuito de explorar o comportamento passado e também de prever o comportamento futuro de um determinado problema. Essa análise desempenha um papel fundamental em diversas áreas do conhecimento, trazendo uma série de benefícios e aplicações práticas, como a identificação de padrões e tendências, previsões, tomadas de decisões estratégicas, planejamento e controle, otimização de recursos entre outros. É neste contexto que implementaremos as Séries Temporais com focos em queimadas X desmatamentos na região da Amazônia Legal que possui uma área total de 5,2 milhões de quilômetros quadrados e abrange 9 estados brasileiros: Acre, Amapá, Amazonas, Mato Grosso, Pará, Rondônia, Roraima , Tocantins e parte do estado do Maranhão, e que atualmente vem se agravando cada vez mais. Com a utilização de Séries Temporais e Machine Learning , o objetivo principal desse projeto é contribuir com a diminuição dos focos de queimadas e desmatamentos, os quais tem um grande significado de impacto sobre o clima, altos níveis de poluentes atmosféricos que podem causar danos à saúde humana e dos animais, destruição dos recursos florestais, biodiversidade e propiciar um estudo e análise mais aprimorada para que as predições e sazonalidades nos ajudem a tomarmos decisões mais diretivas e assertivas. Outro ponto importante é analisar onde ocorrem os maiores focos e detectar as anomalias para identificarmos os comportamentos incomuns desses eventos. Para fazer essas previsões até 2025, utilizaremos as informações da base de dados do TerraBrasilis de 2019 a 2024 e, também a base de dados do INMET(Instituto Nacional de Meteorologia) que possuem variáveis como: precipitação total, temperatura máxima , umidade relativa máxima e vento(velocidade), para que nos ajude a desenvolver modelos de previsões mais precisos e robustos e que possam nos ajudar na gestão e na prevenção de incêndios florestais.

Palavras-chave: Séries Temporais, comportamento futuro, Aprendizado de Máquina, Tendências e Sazonalidades comportamentais.

**Metodologia:**

Nesse projeto serão aplicadas técnicas de análises estatísticas para avaliar importantes propriedades como estacionariedade, sazonalidade , autocorrelação que são chamadas estatísticas descritivas( valores como média, mediana, desvio padrão, mínimo e máximo) entre outras. Já na estatística inferencial, onde procuraremos encontrar tendências que nos permita fazer previsões, avaliar os modelos, testar hipóteses e procurar entender como a temperatura, a umidade, a precipitação e a velocidade dos ventos, podem influenciar ou impactar diretamente nas queimadas e desmatamentos.

Destacaremos também, os modelos SARIMAX e VAR.

O SARIMAX (Seasonal Autoregressive Integrated Moving Average with Exogenous Factors) é um modelo estatístico, que leva em conta a sazonalidade; é a extensão do modelo SARIMA, que permite a inclusão de variáveis exógenas, ou seja, variáveis que podem afetar a Série Temporal que não são parte do próprio processo de Séries Temporais. Essas variáveis já citadas acima, desempenham papéis significativos no contexto queimadas e desmatamentos e que podem influenciar, pois devido a altas temperaturas, a probabilidade de propagação do fogo devido às condições de seca torna a vegetação mais suscetível a incêndios, assim como a umidade, quando elevada, a probabilidade de incêndios é bem menor. Quanto a velocidade dos ventos, podem desempenhar um papel crucial na propagação do fogo quando os ventos são fortes, aumentando a área afetada pelo incêndio e a direção que determina para onde o fogo se espalhará.

A biblioteca necessária para trabalhar com dados em ST, plotagem e modelos estatísticos é “pmdarima”, que fornece uma interface conveniente para ajustar, diagnosticar e fazer previsões com o SARIMAX. Disponível em: <https://www.geeksforgeeks.org/complete-guide-to-sarimax-in-python/> e <https://365datascience.com/tutorials/python-tutorials/sarimax/> Acesso em: 23/04/24

O modelo VAR ( Vector Auto Regressive) é um modelo estatístico utilizado para analisar e fazer previsões em séries temporais multivariadas, ou seja, conjunto de dados que consistem em múltiplas séries temporais observadas simultaneamente ao longo do tempo, como neste projeto, onde vamos prever focos de queimadas para 2025 levando em consideração as temperaturas, velocidade dos ventos, chuvas e outros.

O modelo VAR é definido como:

*Yt*​=*c*+*A*1​*Yt*−1​+*A*2​*Yt*−2​+…+*Ap*​*Yt*−*p*​+*ϵt*​ onde,

*Yt* é um vetor de dimensão p contendo variáveis no tempo t

*c* é uma matriz de constantes

*A*1, *A*2... *Ap*​ são matrizes de coeficientes auto-regressivos

*ϵt*​ é um vetor de resíduos ou erros

Fonte: Elaborado pelo próprio autor

Resumindo, o modelo VAR é uma ferramenta flexível e poderosa para análise e previsão de séries temporais multivariadas, permitindo capturar relações de retroalimentação e dinâmicas complexas entre múltiplas variáveis ao longo do tempo.

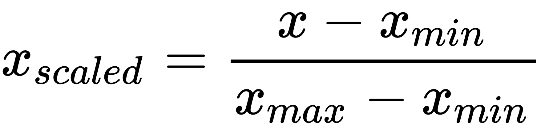
Logo, no projeto, o VAR irá analisar e descrever a interação entre os focos de queimadas com as temperaturas, por exemplo. Uma vez estimado esse modelo, poderemos nos perguntar se existe ou não causalidade nessa relação ou em relação às outras variáveis. Disponível em: <https://shallbd.com/pt/entendendo-a-formula-do-modelo-var-um-guia-abrangente/> e <https://1library.org/article/modelo-autorregressivo-vetorial-var-m%C3%A9todos-de-an%C3%A1lise.qmj1g69q> Acesso em: 23/04/24.

O StandardScaler é uma técnica comum de pré-processamento de dados utilizada em Machine Learning e estatística. Ele é parte do pacote scikit-learn em Python e é usado para padronizar os recursos de um conjunto de dados, o que significa centralizar os dados em torno de zero e dimensioná-los pela variância. Ou seja, transforma os dados de forma que a média de cada recurso seja zero e o desvio padrão seja 1. Isso é importante pois muitos algoritmos em Machine Learning assumem que os dados estão em uma escala similar e têm distribuição normal. Padronizar os dados ajuda a garantir que eles atendam a essas suposições, o que geralmente leva a um melhor desempenho dos modelos.

Foi exatamente o que fizemos no projeto. Os dados do TerraBrasilis eram diários e tivemos que transformá-los em mensais . O mesmo com os dados do INMET, e ainda, os valores foram arredondados para duas casas decimais. E os dados foram normalizados e separados em treino e teste para iniciarmos as previsões.

Já o MInMaxScaler é uma ferramenta do pacote sklearn-preprocessing do Python utilizada para normalizar os dados. A normalização é uma técnica de pré-processamento de dados que ajusta a escala dos valores numéricos, garantindo que todos os valores fiquem dentro de um intervalo específico, geralmente entre 0 e 1. Isso pode melhorar o desempenho de muitos algoritmos de Machine Learning que são sensíveis à escala dos dados.

O MinMaxScaler transforma os valores de um conjunto de dados de acordo com a seguinte fórmula:



Fonte:https://www.oreilly.com/api/v2/epubs/9781788627306/files/assets/ffb3ac78-fd6f-4340-aa92-cde8ae0322d6.png. Acesso em 23/04/24.

Onde:

*X*  é o valor original;

*X*min é o menor valor do conjunto de dados;

*X*max é o maior valor do conjunto de dados;

*X*scaled é o valor normalizado.

Fonte de dados:

O banco de dados, obtido pelo portal TerraBrasilis, plataforma desenvolvida pelo INPE para organização, acesso, consulta, análise de dados geográficos do Instituto PRODES e DETER que qualifica e espacializa as ocorrências de focos de queimadas X desmatamentos por estado brasileiro.

O INPE, visando o aprimoramento contínuo de seus produtos através das pesquisas realizadas, e atendendo as demandas do Governo Federal lançou um novo painel de informações que relaciona as localizações dos focos de calor em vegetação mapeadas pelo Projeto Queimadas e dos dados de desmatamento produzidos pelos projetos PRODES e DETER para a região da Amazônia Legal Brasileira. Tal produto considera a relação entre o processo de fogo em vegetação e desmatamento no portal TerraBrasilis.

São considerados desmatamentos recentes aqueles mapeados pelo PRODES nos últimos dois anos de referência. Já o desmatamento consolidado consideram áreas desmatadas e medidas pelo PRODES em um período anterior aos dois anos referência do PRODES. Na data atual, estes correspondem aos polígonos de desmatamento do PRODES mapeados entre 1988 e 2017. O painel apresenta o número total de focos de calor em cada uma dessas classes.

Para entender melhor , o TerraBrasilis opera os projetos de monitoramento da Floresta Amazônica Brasileira por satélite (PRODES) e o Sistema de Detecção de Desmatamento em Tempo Real (DETER), que utilizam imagens de sensoriamento remoto para acompanhar alterações na cobertura florestal na região. Foi desenvolvido com a missão de melhorar e aprimorar continuamente as suas ferramentas de suporte que possibilita a visualização sinótica de indicadores de áreas críticas do desmatamento. Tem, também como principal objetivo, subsidiar o planejamento de ações de fiscalização pelas instituições responsáveis, tanto a nível federal quanto estadual, e que possa ser útil para que outros setores da sociedade possam entender os múltiplos processos, como também, discutir e demandar ações efetivas para controle desses processos na região.

O código usado na implementação específica do MAS é baseado em bibliotecas abertas e livres de licença e está aberto, possibilitando a colaboração de terceiros. Disponível em : [http://terrabrasilis.dpi.inpe.br/app/dashboard/fires/biomes/aggregated/#](http://terrabrasilis.dpi.inpe.br/app/dashboard/fires/biomes/aggregated/)

Será utilizada também, a base de dados do INMET( Instituto Nacional de Meteorologia) responsável por diversas atribuições relacionadas à meteorologia e climatologia no Brasil, como monitoramento e alertas meteorológicos, previsão do tempo, estudos climatológicos entre outros, para a promoção da segurança e bem-estar da população brasileira, bem como no apoio a diversas atividades econômicas e sociais que dependem das condições meteorológicas. Disponível em: <https://tempo.inmet.gov.br/TabelaEstacoes/A001ml> Acesso em 07/04/24.

Será usada a linguagem de programação em Python e as bibliotecas mais importantes, além do pandas e numpy, serão: seaborn, matplotlib, scipy, statsmodels, e outras que no decorrer do projeto utilizaremos.

**Introdução:**

As queimadas e desmatamentos na Amazônia Legal brasileira aumentaram drasticamente nos últimos anos, gerando recordes.



Disponível em:<https://www.ecycle.com.br/metodologia-inedita-mapeia-queimadas-no-cerrado-ao-longo-de-20-anos/>. Acesso em: 20/03/24

Quando florestas e outras áreas verdes são queimadas, mesmo que fiquem a quilômetros de distância, liberam enormes quantidades de gazes poluentes que afetam à saúde e podem contribuir para aquecer a Terra. E se grandes áreas com árvores são derrubadas, perdemos o serviço que elas prestam a todos os seres vivos, como: filtrar o ar, produzir umidade, regular a temperatura, produzir alimentos entre outros. Disponível em: <https://criancaenatureza.org.br/pt/clima/queimadas-desmatamento/>. Acesso em 20/03/24

A Amazônia Legal abrange uma área de 5,2 milhões de quilômetros quadrados e tendo em vista a crescente degradação do meio ambiente, torna-se necessário o desenvolvimento de monitoramento mais intensivo e mais eficiente nessas áreas. E para esse monitoramento usar tecnologias inovadoras certamente irão auxiliar na construção e utilização de soluções para lidar com essa complexidade do bioma.

Segundo a Revista Galileu de fevereiro de 2024, a área queimada no Brasil aumentou em 248% em janeiro de 2024. Isso representa 287 mil hectares queimados e o MapBiomas diz que o bioma mais afetado foi a Amazônia. Disponível em:<https://revistagalileu.globo.com/ciencia/meio-ambiente/noticia/2024/02/area-queimada-no-brasil-aumentou-248percent-em-janeiro-de-2024-estima-pesquisa.ghtml> Acesso em 21/03/24

Mas em compensação, o desmatamento que foi intenso em 2023, agora em 2024, segundo o portal O Tempo, cai em 41,7% no primeiro trimestre de 2024, dados do Sistema DETER do INPE. Disponível em:<https://www.otempo.com.br/brasil/desmatamento-na-amazonia-cai-42-no-primeiro-trimestre-de-2024-1.3361036> Acesso em: 21/03/24

O Governo Federal abriu adesão as cidades que queiram participar do Programa União com Municípios pela Redução de Desmatamento e Incêndios Florestais, iniciativa lançada em setembro de 2023, onde o Ministro do Meio Ambiente e Mudança do Clima anunciou um investimento de 600 milhões do Fundo Amazônia para 69 municípios que atuem no controle e monitoramento, recuperação da vegetação nativa entre outros. Disponível em:<https://agenciabrasil.ebc.com.br/geral/noticia/2024-04/prazo-para-aderir-ao-controle-do-desmatamento-comeca-dia-12-de-abril> Acesso em: 21/03/24

Logo o nosso projeto pretende ajudar a minimizar esse problema fazendo previsões até 2025, utilizando as informações do banco de dados do TerraBrasilis , juntamente com banco de dados do INMET, realizar a detecção de anomalias podendo assim saber quais os estados que tem mais ou menos incidências do evento, e também encontrarmos a sazonalidade, ou seja, saber qual período de maior incidência de queimadas e desmatamento e procurar saber se alguma variável exógena pode ou não influenciar nas queimadas e desmatamentos e assim podermos tomar decisões mais precisas e seguras. Com todas essas informações podemos ajudar cada município a prevenir e preservar o bioma mantendo a qualidade do ar e a saúde de muitas pessoas e animais.

A relevância do problema das queimadas e desmatamentos está vinculada aos Objetivos de Desenvolvimento Sustentável no que tange a saúde e bem estar, ação climática e vida terrestre. Deste modo, com a implementação das Séries Temporais podemos colaborar efetivamente com a preservação dos ecossistemas terrestres, das florestas e da biodiversidade.

**Referencial Teórico:**

De acordo com o acontecimento em 10 de agosto de 2019 em Rondônia, conhecido como “Dia do Fogo” devido a pior onda de incêndio da Amazônia, Jean Cardoso decidiu trabalhar com os dados disponibilizados pelo INPE. Ele fez a compreensão dos dados, usou a estatística descritiva, observou se a série era ou não estacionária, procurou por sazonalidade e tendência. Utilizou a técnica de Hold-Out, método simples para avaliar um classificador, que tem a função de atribuir itens de dados em uma determinada categoria ou classe de destino. Treinou três diferentes modelos de previsão: Suavização Exponencial, baseado na descrição da tendência e da sazonalidade, o ARIMA que é uma das principais abordagens usadas na previsão de Séries Temporais e Redes Neurais que permite um relacionamento não linear complexo entre a variável resposta e seus preditores. A linguagem utilizada foi o software R.Como resultado dessa análise, Jean observou que a série apresentava uma sazonalidade muito forte, mas constante, enquanto a tendência demonstrou sofrer muitas alterações e que analisando os gráficos, as piores épocas de queimadas são os meses de julho a agosto e, no mês de setembro foi o mês mais crítico por causa da chegada da estação seca. Diz que fica complicado compreender a tendência pois o foco de queimadas da Amazônia é influenciado por muitos fatores externos. Dos três modelos aplicados, o ARIMA foi o de melhor resultado e diz que os modelos como ARIMAX ou uma Regressão Dinâmica seriam exemplos a serem explorados em trabalhos futuros. Artigo publicado em 25 de setembro de 2019 – “Pray For Amazonia – uma breve análise de Séries Temporais, realizada pelo PhD Jean Cardoso.Disponível em:<https://www.linkedin.com/pulse/prayforamazonia-uma-breve-an%C3%A1lise-de-s%C3%A9ries-temporais-cardoso/> Acesso: 07 de março de 2024.

Este artigo refere-se a uma pesquisa de crescimento de número de automóveis no município de Joinville/SC e por meio da análise de séries temporais podem auxiliar os setores públicos e o setor automobilístico a controlar esse crescimento desenfreado. Através de alguns métodos como os autorregressivos integrados de média móvel (ARIMA), modelos de suavização exponencial e combinação de modelos e previsões de dados mensais de janeiro de 2003 a março de 2014, pode-se observar que os modelos autorregressivos integrados de média móvel apresentaram os menores erros, sendo então o modelo SARIMA o mais adequado para tais previsões. O principal objetivo dessa análise é auxiliar o poder público em tomadas de decisões relacionada à infraestrutura, planejamento de tráfego e definição de políticas públicas devido à expansão da frota. Quanto a avaliação dos modelos foram realizadas em duas etapas: a primeira no que tange ao diagnóstico do modelo( se o modelo atende as condições estabelecidas) e a segunda é a análise dos erros de previsão. A série é não estacionária, a tendência é crescente e existe uma sazonalidade (mostrada no gráfico). Como conclusão, o modelo ARIMA foi o qual representou melhor a série e que para trabalhos futuros considera como sugestão o estudo considerando as variáveis causais e propõe modelos econométricos para realizar previsões e verificar o menor erro. A linguagem de software utilizada foi o R por meio do pacote forecast. Revista Espacios – ano 2016, página 29- “Análise de séries temporais para previsão da evolução do número de automóveis no município de Joinville” Disponível em:<https://qualimetria.ufsc.br/files/2016/05/Revista-ESPACIOS-_-Vol.pdf> Acesso em: 22 de abril de 2024.

Este artigo refere-se sobre a importância das florestas pois regulam o equilíbrio do ecossistema. Na Indonésia os incêndios florestais ocupam o segundo lugar, sendo o primeiro, o desmatamento, e podem ocorrer devido à fatores humanos e fatores naturais. O estudo feito tem como objetivo prever o número de ocorrências de focos de incêndio e os métodos utilizados para fazer tais previsões são o ARIMA (Autoregressive Integrated Moving Average) e o SARIMA (Seasonal Autoregressive Integrated Moving Average). Obtiveram como resultados de avaliação de RMSE (Root Mean Square Error) para o ARIMA igual a 6,61, enquanto para o SARIMA igual a 7,61. Para o MSE (Erro Quadrático Médio) para o ARIMA igual a 43,7 e para o SARIMA igual a 58,05. Com base nesses resultados, pode-se concluir que o ARIMA apresentou excelente desempenho e precisão na descrição da tendência. Jornal SISFOKOM –Vol.13, publicado em fevereiro de 2024.-“Predicting the Number of Forest na Land Fire Hostpot Occurrences Using the ARMA and SARIMA Methods”. Disponível em:[https://jurnal.atmaluhur.ac.id/index.php/sisfokom/article/view/2018/978 Acesso em:25](https://jurnal.atmaluhur.ac.id/index.php/sisfokom/article/view/2018/978%20%20%20Acesso%20em:25) de abril de 2024

Neste artigo leva-se em conta desde dados de incêndios anteriores até fatores climáticos e aborda a complexidade dos incêndios florestais e seus impactos. Com a capacidade de prever emissões mensais de incêndios com até seis meses de antecedência, oferece uma oportunidade significativa para melhorar a gestão de incêndios e reduzir seus impactos negativos. Ao utilizar um método estatístico como o ARIMAX e otimizar seus parâmetros para mais de 1300 regiões diferentes, o sistema demonstra uma abordagem abrangente e adaptável. A validação cruzada do sistema, considerando diferentes combinações de preditores e prazos de previsão, é crucial para garantir sua confiabilidade e precisão. E também, o sistema é capaz de explicar uma parte significativa da variabilidade nas emissões globais de incêndios e de resolver padrões espaciais detalhados de anomalias de emissões em regiões com atividades significativas de incêndio florestais. Ferramenta valiosa para governos, agências de gestão de emergências e comunidades afetadas na preservação e resposta a incêndio florestais. Como conclusão, em geral, os modelos utilizados baseados em ARIMAX produziram melhores previsões do que os modelos mais simples baseados na persistência de anomalias ou na climatologia das emissões de incêndios anteriores. O ARIMAX foi capaz de lidar com diferentes fontes de previsão de incêndios florestais em prazos de 1 a 6 meses. JAMES Journal of Advances in Modeling Earth Systems, publicado em agosto de 2020, “Previsão de emissões globais de incêndios em escalas de tempo subsazonais a sazonais. Disponível em:<https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2019MS001955> Acesso em 25 de abril de 2024

Em relação ao nosso projeto, os estudos acima citados nos ajudam a ter mais ideias de implementação, de explorar melhor os dados para obter previsões mais precisas, de saber o por quê em períodos específicos há mais ou menos queimadas e desmatamentos, refletir melhor sobre sazonalidades e tendências, enfim, são fundamentais para a legitimação de pesquisas e estudos que contribuem com a sociedade. Sabemos que as queimadas e desmatamentos são pontos difíceis de se resolver, mas nesse projeto iremos procurar fazer o melhor para ajudar o planeta, melhorar as condições atmosféricas, assim como a saúde dos seres humanos, a fauna e a flora. Queremos também deixar claro que inicialmente houve dificuldade em encontrarmos referenciais teóricos que correlacionassem métodos ao problema proposto ao nosso projeto. Na próxima etapa, iremos desenvolver outro modelo para que possamos compará-los e analisarmos qual terá melhores resultados.

**Diagrama de Solução:**

**EDA E PRÉ-PROCESSAMENTO DOS DADOS: estão descritos nos próprios códigos**

A análise exploratória é uma etapa essencial em projetos de Ciências de Dados, pois busca compreender melhor os dados, seja por identificar padrões, tendências, sazonalidades, outliers, possíveis relações entre as variáveis, entre outros. Explorar informações relevantes para orientar respostas aos objetivos específicos.

Ao iniciar o projeto foi necessário instalar pacotes, importar bibliotecas que possuem funções específicas a serem utilizadas nas linhas de código, como mostra os quadros abaixo:

# instalar pacotes adicionais:

|  |  |
| --- | --- |
| !pip install pycaret | biblioteca de Machine Learning |
| !pip install pmdarima | estimador ARIMA para Python |

Fonte; Elaborado pelo próprio autor.

# importar as bibliotecas necessárias:

|  |  |
| --- | --- |
| import numpy as np | manipulação de arrays |
| import matplotlib.pyplot as plt | visualização de dados |
| import seaborn as sns | visualização estatística dos dados |
| import datetime as dt | manipulação de datas e horários |
| import plot.express as px | criar visualizações |
| import os | interação com o sistema operacional |
| import requests | fazer requisições HTTP em python |
| import statsmodels.tsa.statespace.varmax import VARMAX | método de previsão VARMAX |
| from sklearn.preprocessing import StandardScaler | modelo estatístico para análises e previsões |
| from sklearn.metrics import mean\_square\_error, mean\_absolute\_error | métricas de avaliação |
| from statsmodels.tsa.stattools import adfuller | teste adf (séries estacionárias) |
| from statsmodels.tsa.seasonal import STL | decompor série temporal |
| import statsmodels.api as sm | importar o pacote statsmodels |
| from statsmodels.tsa.statespace.sarimax import SARIMAX | método de previsão SARIMAX |
| from pmdarima.arima.stationarity import ADFTest,KPSSTest,PPTest | testes de estacionariedade em séries temporais |
| from pmdarima.arima.utils import ndiffs | útil para determinar o nº de diferenciações para tornar uma ST estacionária |
| from pmdarima.model\_selection import train\_test\_split | dividir os dados em treino e teste para avaliar o desempenho do modelo |
| from statsmodels.tsa.api import VAR | modelo de previsão VAR |
| from statsmodels.graphics.tsaplots import plot\_acf, plot\_pacf | plotar gráficos de autocorrelação e autocorrelação parcial |

Fonte: Elaborado pelo próprio autor.

Primeiramente foi realizada a leitura dos dados do TerraBrasilis, feita uma visão geral como a quantidade de registros e atributos do dataset, valores ausentes e duplicados, formatação da data, estatística descritiva (média, mediana, desvio padrão, valor mínimo e máximo, quartis, entre outros), observar a menor e maior data, aplicar a função para saber os tipos de dados, plotar um gráfico de dispersão simples para verificar os focos em cada estado, renomeamos as variáveis, feita a média móvel simples e seu plot, entre outras análises contidas no Google Colaboratory. Para a variável numérica “focuses” foi criado um gráfico do tipo histograma para a visualização e compreensão da distribuição dos focos de queimadas e desmatamentos.

Foram plotados outros gráficos para visualização, como de ocorrências por estado e de ocorrências por classe de cada estado.

No início da análise dos dados, sentimos a necessidade de mais uma base , já que o TerraBrasilis só apresenta uma coluna numérica. Decidimos, então, realizar a junção de outro dataset, o INMET que nos traz informações sobre temperatura, velocidade e direção dos ventos, umidade, nebulosidade, chuva entre outros. Isso foi importante para que nossa análise se tornasse mais minuciosa no que tange a direcionar outras variáveis , ou variáveis exógenas que podem impactar ou não no aumento do número de focos de queimadas e desmatamentos. Sabemos que a maioria dessas ocorrências são causadas por ações humanas. Mas até que ponto isso acontece?

É isso que iremos analisar e descobrir aplicando as estatísticas e os modelos em Séries Temporais.

Como ambos datasets se tratavam de valores diários, achamos melhor converter os dados para mensais usando o “datetime” e em seguida, converter a coluna “date” como índice novamente.

PRÉ-PROCESSAMENTO DOS DADOS:

O pré-processamento dos dados é um conjunto de atividades que envolvem preparação, organização e estruturação dos dados. Uma etapa fundamental que precede a realização de análises e predições, determina a qualidade final dos dados que serão analisados e pode, inclusive, impactar nos modelos de previsão.

Iniciamos com a normalização e preparação dos dados para atender os requisitos das bibliotecas, checagem da série (estacionária ou não estacionária) , divisão do conjunto em treino, teste para validação e fazer as previsões utilizando os modelos SARIMAX e VAR (detalhados na metodologia). Mas antes disso, foi necessário

**MODELO BASE: está descrito no próprio código**

**RESULTADOS:**

O objetivo do projeto é contribuir com a diminuição dos focos de queimadas e desmatamentos da região da Amazônia Legal e propiciar um estudo e uma análise mais aprimorada usando Séries Temporais e Machine Learning.

Primeiramente, vamos citar os gráficos que iniciam após a leitura dos dados do INMET.

Criamos gráficos interativos para visualizar o comportamento das variáveis independentes ao longo do tempo, em todos os estados da Amazônia Legal. Demonstra o comportamento geral de janeiro de 2019 a janeiro de 2024 em todos os meses.

Em primeiro momento aparecem várias linhas amontoadas, mas como se trata de um gráfico interativo onde ao clicarmos em cada variável na legenda, essa linha desaparece, podendo assim ter uma visualização individual de cada variável, juntamente com seus valores das médias mensais ao passar o mouse sobre a linha. São, portanto, gráficos auto-explicativos e qualquer pessoa pode entender o que eles pretendem demonstrar.

Fizemos outro gráfico interativo, logo após a leitura dos dados do Portal TerraBrasilis, onde aparece no gráfico a variável dependente “focos”, por isso esse gráfico possui somente uma única linha. Aqui estamos detalhando o número de focos de queimadas para cada estado. Ao passar o mouse sobre a linha podemos ler as informações como mês, ano e o número de focos dos eventos. A partir desse gráfico podemos notar a sazonalidade entre os focos de queimadas, geralmente em junho a dezembro, pois a Amazônia experimenta uma estação seca no período de julho podendo ser estendido até dezembro. Durante esse período a precipitação diminui, reduzindo a umidade do solo e da vegetação, o que facilita a ignição e a propagação de incêndios.

Agricultores e pecuaristas, muitas vezes usam o fogo como uma ferramenta para limpar a terra para o plantio ou pastagem, método conhecido como “queima controlada”, e é mais comum durante a estação seca quando a vegetação está seca e queima mais facilmente.

E fizemos, também, a Matriz de Correlação para cada estado da Amazônia Legal, ou seja, nos mostra a correlação entre as variáveis. A cor amarela representa alta correlação, enquanto a cor azul representa baixa correlação entre tais variáveis.

Agora vamos demonstrar os resultados do Modelo Base**, o Modelo SARIMAX**.

**No Forecast para o estado do Acre**- temos um período de sazonalidade entre julho a novembro.Os maiores focos aparecem no mês de setembro e os menores entre dezembro e abril. Os dados de números de focos de queimadas para janeiro de 2024 são 255 focos e para janeiro de 2025 foi previsto para 500 focos. Quanto as métricas de erro estão indicando um erro significativo para as previsões. Talvez a inclusão de outras variáveis exógenas ou ajustes nos parâmetros do modelo, possam ajudar a reduzir esses erros e melhorar as previsões.

**Forecast do estado do Amapá** – período de sazonalidade entre agosto a dezembro. Os maiores focos aparecem no mês de outubro e os menores entre janeiro e julho. Para 2024, o número de focos é de 3 e para janeiro de 2025 é de 47. Quanto as métricas de erro indicam erros médios, menores em comparação ao estado do Acre, mas com espaço para melhorias como inclusão de mais variáveis exógenas, ajuste nos parâmetros do modelo para que possam reduzir os erros e melhorar as previsões.

**Forecast do estado do Amazonas** – período de sazonalidade entre julho a novembro. Os maiores focos aparecem no mês de agosto e os menores entre janeiro a maio. Para 2024, temos 160 focos e para janeiro de 2025 temos 732 focos de queimadas. Quanto as métricas de erro, o modelo enfrenta desafios significativos na previsão, com erros médios bastante elevados. Isso pode ser devido a diversas variáveis complexas que influenciam os focos de queimadas na região. Além das mesmas sugestões para melhoria, sugerimos , também, a utilização de modelos alternativos para captar melhor as dinâmicas específicas do Amazonas.

**Forecast do estado do Maranhão** – período de sazonalidade entre junho a janeiro. Os maiores focos aparecem no mês de novembro e os menores entre março e abril. Para 2024 temos 131 focos e para janeiro de 2025 temos 227 focos de queimadas. Quanto as métricas de erro indicam que o modelo tem um desempenho razoalvelmente bom na previsão, especialmente se comparado com o Amazonas. A baixa MSE e RMSE indicam que o modelo consegue prever os focos com uma precisão aceitável para a região. As melhorias no modelo ainda podem ser consideradas para reduzir ainda mais os erros.

**Forecast do estado do Mato Grosso** – período de sazonalidade entre julho a novembro. Os maiores focos aparecem no mês de setembro e os menores em dezembro. Para 2024 temos 389 focos e para janeiro de 2025 temos 327 focos de queimadas. Quanto as métricas de erro indicam que o modelo teve um desempenho pior na previsão de focos de queimadas no estado, especialmente quando comparado ao estado do Maranhão. A alta MSE e RMSE indicam que o modelo tem dificuldades em prever os focos de queimadas com precisão. Para melhorias sugerimos revisar os dados de entrada, explorar novas variáveis, testar diferentes tipos de modelos e ajustar hiperparâmetros.

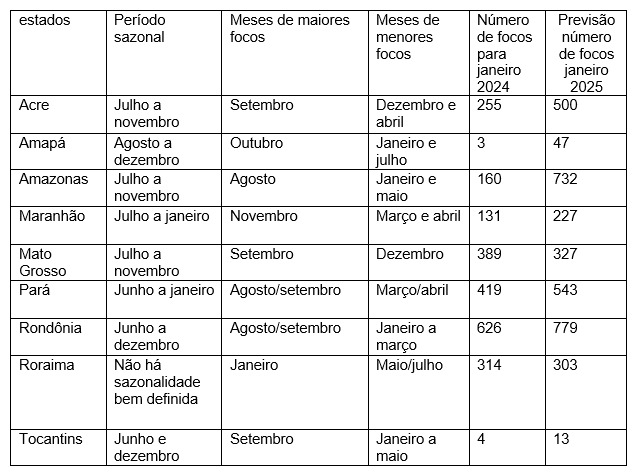
**Forecast do estado do Pará** – período de sazonalidade entre junho a janeiro. Os maiores focos entre os meses de agosto e setembro e os menores entre março e abril.Para janeiro de 2024 temos 419 focos e para janeiro de 2025 temos 543 focos de queimadas. Quanto as métricas de erro indicam que o modelo tem um desempenho insatisfatório na previsão de focos de queimadas no estado. A MSE e a RMSE mostram que o modelo tem dificuldades na previsão com precisão. Para melhorias sugerimos a verificação dos dados de treinamento, explorar novas variáveis, testar diferentes modelos, busca mais abrangente de hiperparâmetros, considerar fatores sazonais específicos que podem estar influenciando os focos de queimadas.

**Forecast do estado de Rondônia** – período de sazonalidade entre junho a dezembro. Os maiores focos são entre os meses de agosto e setembro e os menores entre os meses de janeiro a março. Para janeiro de 2024 temos 626 focos e para janeiro de 2025 temos 779 focos de queimadas. Quanto as métricas de erro indicam que o modelo tem um desempenho moderado na previsão no estado. Embora os erros sejam ainda significativos, eles são menores comparados ao estado do Pará, mostrando que há um potencial de melhorias específicas para essa região. As possíveis melhorias seriam verificar se os dados usados para treinamento estão corretos e se há outliers que precisam ser tratados, explorar novas variáveis ou transformar as existentes para que possam melhorar o desempenho do modelo, testar diferentes tipos de modelos baseados em Machine Learning ou Deep Learning, ajustar os hiperparâmetros e incluir fatores externos, como políticas de desmatamento, ações governamentais e mudanças climáticas que possam impactar os focos de queimadas.

**Forecast do estado de Roaraima** - Neste estado não há uma sazonalidade bem definida, mas o mês de maior foco é janeiro e os de menor focos são maio a julho. Para janeiro de 2024 temos 314 focos e janeiro de 2025 temos 303 focos de queimadas. As métricas indicam que o modelo tem um desempenho relativamente bom na previsão de focos de queimadas no estado, com erros menores comparados a outros estados da Amazônia Legal. Logo, as ações para melhorias seriam praticamente as mesmas ( anteriores).

**Forecast do estado do Tocantins** – o período de sazonalidade está entre junho e dezembro.Os maiores focos está no mês de setembro e os menores entre os meses de janeiro a maio. Para 2024 temos 4 focos e para janeiro de 2025, prevemos 13 focos de queimadas. Para as métricas de erro, Tocantins apresenta algumas das menores métricas , sugerindo que o modelo possui um bom desempenho de previsão para este estado. Comparativamente, Tocantins tem o menor RMSE, indicando previsões mais precisas.

Para facilitar a visualização, elaboramos uma tabela com todas as comparações acima.



- Conclusão para o forecast SARIMAX:

Os resultados mostraram que há uma variação no desempenho do modelo SARIMAX entre os diferentes estados. Tocantins e Maranhão apresentaram métricas de erro significativamente menores, indicando previsões mais precisas. Por outro lado, Pará e Mato Grosso mostraram erros muito maiores, sugerindo que as previsões para esses estados são menos precisas. As possíveis causas dessas variações são: a qualidade dos dados disponíveis para cada estado, pois podem influenciar os resultados. Variáveis meteorológicas como precipitação e temperatura podem ter relações diferentes com os focos de queimadas em diferentes regiões. Fatores regionais específicos, como políticas de desmatamento, práticas agrícolas e controle de queimadas, podem afetar os resultados.

- Conclusão das métricas de erro (SARIMAX)

O desempenho do modelo SARIMAX na previsão de focos de queimadas varia significativamente entre os estados da Amazonia Legal. Enquanto Tocantins e Roraima apresentam métricas de erro mais baixas e, portanto, previsões mais precisas. Estados como Acre, Amazonas e Mato Grosso enfrentam desafios significativos. Ajustes nos parâmetros do modelo, inclusão de variáveis exógenas adicionais e a exploração de modelos alternativos são estratégias que recomendamos para melhorar a precisão das previsões em estados com desempenho limitado e insatisfatório.

Aqui demonstramos os resultados do **Modelo VAR.**

Primeiramente, vamos analisar os resultados dos Testes de Dickey-Fuller Aumentado (ADF) e das Métricas de Erro. O Teste ADF é utilizado para verificar se uma série temporal é estacionária. A estacionariedade é uma propriedade essencial para a modelagemde séries temporais.

Antes da diferenciação, os resultados para a váriável dependente “focos” e para as variáveis independentes “prepicitação total, temperatura máxima, umidade relativa máxima e velocidade dos ventos”, juntamente com o p-valor alto, ou seja, maior do que 0,05, nos indicava que a série era não estacionária. A partir daí, realizamos a diferenciação dos valores através da interpolação e obtivemos o p-valor muito menor do que 0,05, indicando uma série estacionária, o que é positivo e necessário para a modelagem correta das séries temporais.

Quanto as métricas de erro, segue abaixo o descritivo por estado:

Acre – indicam que o modelo VAR tem um desempenho limitado na previsão nas ST de focos de queimadas e precipitação no Acre. Os valores elevados das métricas sugerem que há espaço para melhorias como: ajustar os parâmetros do modelo, considerar variáveis exógenas adicionais ou até mesmo explorar outros modelos de previsão entre outros.

Amapá – indicam que o modelo VAR tem um desempenho moderado na previsão. Os valores elevados de MSE e RMSE sugerem a presença de outliers e para melhorias sugerimos ajuste nos parâmetros do modelo, a consideração de variáveis adicionais, exploração de outros modelos, enfim, as mesmas melhorias anteriores.

Amazonas – indicam que o modelo VAR tem um desempenho moderado na previsão. Os valores elevados de MSE e RMSE sugerem a presença de erros de previsão significativos e, a diferença entre MAE e RMSE indica a presença de outliers. Para melhorias, seguem as mesmas sugestões acima (estado Amapá).

Maranhão – indicam que o modelo VAR tem um desempenho razoável na previsão. No entanto, há espaços para melhorias como: ajustar os parâmetros, podem ajudar na precisão das previsões.

Mato Grosso – embora o modelo VAR seja capaz de fornecer previsões razoáveis para o estado, as métricas de erro indicam que há margem para melhorias, como: uma análise mais profunda dos dados, pois em comparação com os outros estados da Amazônia Legal, Mato Grosso apresenta métricas de erros muito altas.

Pará – apresenta um desempenho moderado, também com sugestão para melhorias.

Rondônia – o modelo VAR oferece uma abordagem promissora para a previsão nesse estado, pois em comparação com os outros estados, as métricas de erro estão numa faixa intermediária, ainda sugerindo melhorias.

Roraima – indicam que o modelo VAR tem um desempenho satisfatório na previsão. Em comparação com os outros estados, Roraima apresenta métricas de erro relativamente baixas, mas ainda sugerimos melhorias, pois é importante continuar refinando o modelo para garantir previsões mais precisas.

Tocantins – indicam que o modelo VAR tem um desempenho bastante satisfatório na previsão. Em comparação com os outros estados, Tocantins apresenta métricas de erro relativamente baixas, ou seja, o VAR está capturando adequadamente os padrões nos dados e produzindo previsões precisas.

- Conclusão das métricas de erro: (VAR)

O modelo VAR apresentou variações significativas no desempenho entre os diferentes estados da Amazônia Legal. Em alguns estados, o desempenho é moderado a satisfatório, enquanto em outros, há indicações claras de que há espaço para melhorias consideráveis.

Acre, Amapá, Amazonas, Pará e Mato Grosso, nestes estados, o modelo VAR teve um desempenho limitado a moderado. As métricas de erro (MSE e RMSE) indicam que as previsões contém erros significativos e a presença de outliers. Isso sugere que o modelo pode não estar capturando adequadamente as dinâmicas subjacentes dos dados.

Recomendamos ajustar os parâmetros do modelo, considerar variáveis exógenas adicionais e explorar outros modelos de previsão, como modelo de Machine Learning, para melhorar a precisão das precisões.

O desempenho do modelo VAR em Rondônia é intermediário, sendo uma abordagem promissora em comparação com os outros estados. No entanto, ainda há espaço para melhorias, como ajustes finos no modelo e incorporação de variáveis adicionais que podem, potencialmente, melhorar ainda mais as previsões.

Já Roraima e Tocantins apresentaram um desempenho relativamente satisfatório. As métricas de erro são baixas em comparação com outros estados, indicando que o modelo está capturando bem os padrões nos dados e fornecendo previsões razoavelmente precisas. Mas apesar do desempenho satisfatório, recomendamos continuar refinando o modelo para garantir previsões ainda mais precisas.

Executamos o Teste de Causalidade de Granger entre as variáveis dependentes e independentes. A estatística F, que é aquela que indica que a variável independente melhora a capacidade do modelo prever a variável dependente. Quanto mais alto o valor de F , maior a significância da relação e quanto menor o valor de F da variável independente, não contribui para explicar a variabilidade da variável dependente.

Resultados do teste de causalidade de Granger:

**Acre**

foco e precipitação total – há fortes evidências de que a precipitação total pode prever os focos de incêndio.

foco e temperatura máxima – existem algumas evidências de causalidade entre as variáveis, mas não são consistentes.

foco e umidade relativa máxima – há evidências consistentes de que a umidade pode prever os focos de incêndio.

foco e vento,velocidade- hão há nenhuma evidência de causalidade entre as variáveis.

**Amapá**

foco e precipitação total, temperatura máxima e umidade relativa máxima – é um bom preditor dos focos de incêndio até 2 lags.

Foco e vento,velocidade – é um bom preditor dos focos de incêndio até 3 lags.

**Amazonas**

Foco e precipitação total; umidade – mostram evidências de causalidade para os focos de incêndio para vários lags.

Foco e temperatura máxima; vento,velocidade – não mostram evidências de causalidade significativa para os focos de incêndio.

**Maranhão**

foco e precipitação total – indicam ter uma relação de causalidade entre as duas variáveis. Os resultados indicam uma relação significativa entre a precipitação total e os focos de queimadas no Maranhão, sugerindo que a precipitação total é um bom preditor para os focos de queimadas. Esses insights podem ser utilizados para melhorar os modelos de previsão e para desenvolver políticas de prevenção mais eficazes baseadas nas condições climáticas.

Foco e temperatura – para cada número de lag ( de 1 a 5), os resultados mostram valores de F altos e p-valor extremamente baixos, indicando que a temperatura máxima tem influência significativa no foco.

Foco e umidade relativa máxima - para cada número de lag ( de 1 a 5), os resultados mostram valores de F altos e p-valor baixos, indicando que a umidade relativa máxima tem influência significativa em foco, ou seja, pode ser usada para prever os focos de queimadas.

Foco e vento,velocidade – idem acima, indicando também, que para ambas as variáveis existem influências significativas no foco.

**Mato Grosso**

Foco e precipitação total – existe uma evidência significativa de causalidade dos focos sobre a precipitação para os lags 1 a 3, sendo o p-valor menor do que 0,05.

Foco e temperatura máxima – existem evidências significativas, principalmente nos lags 1,3,4 e5 .

Foco e umidade relativa máxima – existem fortes evidências para todos os lags testados.

Foco e vento, velocidade- há evidências de causalidade para os lags 2 a 5.

**Pará**

Foco e precipitação total – indicam fortes evidências de causalidade com significância para todos os lags testados.

Foco e temperatura máxima – a relação de causalidade é menos consistente. Para os lags 2 e 3 há evidências significativas de causalidade, mas não para o lag 1 e, os resultados são mistos para os lags 4 e 5.

Foco e umidade relativa máxima – há evidências robustas de causalidade para todos os lags.

Foco e vento, velocidade – a evidência de causalidade é menos consistente, mas há significância para os lags 2 a 5, mas não para o lag 1.

**Rondônia**

Foco e precipitação total – há evidências fortes de causalidade para todos os lags testados.

Foco e temperatura máxima – há evidências moderadas, especialmente para os lags 2 e 3.

Foco e umidade relativa máxima – há evidências fortes de causalidade para todos os lags testados.

Foco e vento,velocidade- não há evidências significativas de causalidade.

**Roraima**

Foco e precipitação total – existe causalidade entre os focos de queimadas e a precipitação total, especialmente com lags 1 e 2.

Foco e temperatura máxima – não há evidências de causalidade.

Foco e umidade relativa máxima – há causalidade, principalmente para o lag 2.

Foco e vento,velocidade – existe causalidade, notavelmente com lag 2 e 3.

**Tocantins**

Foco e precipitação total – há evidências de causalidade, especialmente os resultados mostram significância estatística para lags 1 a 5.

Foco e temperatura máxima – não há evidências de causalidade. Os resultados mostram alguma significância estatística para os lags 2 a 4, mas não são consistentes.

Foco e umidade relativa máxima – há fortes evidências de causalidade para todos os lags testados.

Foco e vento,velocidade – há fortes evidências de causalidade e a significância estatística é consistente para todos os lags.

Conclusão do Teste de Causalidade de Granger:

* Foco e precipitação total: em muitos estados há evidências significativas de causalidade. Isso sugere que a quantidade de chuva pode influenciar a ocorrência de incêndios florestais em várias regiões. Chuvas intensas podem estimular o crescimento exuberante de vegetação. Embora isso possa inicialmente reduzir o risco de incêndios, quando a vegetação cresce e subsequentemente seca durante períodos de seca, ela pode fornecer uma grande quantidade de “combustível” para incêndios futuros. Em contraste, a falta de chuva pode limitar o crescimento da vegetação, resultando em menos “combustível” disponível para incêndio, mas, no entanto, a vegetação existente pode secar e se tornar altamente inflamável.
* Foco e temperatura máxima: os resultados em relação à temperatura máxima são muito variados. Em alguns estados, como Roraima, não há uma relação clara entre a temperatura e os focos de incêndio, pois os resultados mostram que, com uma defasagem de 3 períodos essa relação não é consistentemente significativa para outros lags. Para os outros estados, a relação pode variar devido a fatores climáticos, geográficos e de uso da terra.
* Foco e umidade relativa máxima: a umidade parece desempenhar um papel importante na ocorrência de incêndios florestais em muitos estados, com evidências de causalidade em várias regiões. A umidade relativa do ar é um fator crítico na ocorrência e propagação de incêndios florestais. Quando a umidade é alta o ar contém mais vapor de água, o que dificulta a ignição e a propagação do fogo. Em contrapartida, quando a umidade é baixa, a vegetação fica mais seca e suscetível a incêndios.
* Foco e vento, velocidade: a velocidade do vento apresenta uma relação de causalidade menos consistente. Em alguns estados há evidências significativas, enquanto em outros não há relação detectada. Ventos fortes podem espalhar chamas e brasas, aumentando a área afetada pelo fogo, mas também podem não estar diretamente relacionados ao surgimento inicial de focos de incêndio.

E por último, fizemos tabelas que demonstram a relação entre as variáveis dependentes (focos) e as demais variáveis com respeito a sazonalidade e em que meses há maiores altas ( último gráfico do modelo VAR).

Acre

foco (F) e precipitação total (P)

|  |  |  |
| --- | --- | --- |
| Período sazonalidade F | Maiores altas P | Relação F e P |
| julho a novembro | Dezembro e fevereiro | ↑ P ↓ F e ↓P ↑F |

foco (F) e temperatura máxima (T)

|  |  |  |
| --- | --- | --- |
| Período sazonalidade F | Maiores altas T | Relação F e T |
| Julho a novembro | Setembro e dezembro | ↑ T ↑ F e ↓ T ↓ F |

foco (F) e umidade relativa máxima (U)

|  |  |  |
| --- | --- | --- |
| Período sazonalidade F | Maiores altas U | Relação F e U |
| Julho a novembro | Março e dezembro | ↑ U ↓ F e ↓U ↑F |

foco (F) e vento (V)

|  |  |  |
| --- | --- | --- |
| Período sazonalidade F | Maiores altas V | Relação F e V |
| Julho e novembro | Setembro e abril | ↑ V ↓ F e ↓V ↑F |

Amapá

foco (F) e precipitação total (P)

|  |  |  |
| --- | --- | --- |
| Período de Sazonalidade F | Maiores altas P | Relação F e P |
| Setembro e Novembro | Abril e Maio | ↑P ↓F e ↓P ↑F |

foco (F) e temperatura máxima (T)

|  |  |  |
| --- | --- | --- |
| Período de Sazonalidade F | Maiores altas T | Relação F e T |
| Setembro e Novembro | Setembro e Outubro | ↑T ↑F e ↓T ↓F |

foco (F) e umidade relativa máxima (U)

|  |  |  |
| --- | --- | --- |
| Período de Sazonalidade F | Maiores altas U | Relação F e U |
| Setembro e Novembro | Fevereiro e Maio | ↑U ↓F e ↓U ↑F |

foco (F) e vento (V)

|  |  |  |
| --- | --- | --- |
| Período de Sazonalidade F | Maiores altas V | Relação F e V |
| Setembro e Novembro | Janeiro e Outubro | ↑V ↓F e ↓V ↑F |

Amazonas

foco (F) e precipitação total (P)

|  |  |  |
| --- | --- | --- |
| Período de Sazonalidade F | Maiores altas P | Relação F e P |
| Julho a novembro | Fevereiro e Abril | ↑F ↓P e ↓P ↑F |

foco (F) e temperatura máxima (T)

|  |  |  |
| --- | --- | --- |
| Período de Sazonalidade F | Maiores altas T | Relação F e T |
| Julho a novembro | Setembro e Outubro | ↑T ↑F e ↓T ↓F |

foco (F) e umidade relativa máxima (U)

|  |  |  |
| --- | --- | --- |
| Período de Sazonalidade F | Maiores altas U | Relação F e U |
| Julho a novembro | Março e Fevereiro | ↑U ↓F e ↓U ↑F |

foco (F) e vento (V)

|  |  |  |
| --- | --- | --- |
| Período de Sazonalidade F | Maiores altas V | Relação F e V |
| Julho a novembro | Janeiro e Outubro | ↓V ↑F e ↑V ↓F |

Maranhão

foco (F) e precipitação total (P)

|  |  |  |
| --- | --- | --- |
| Período de Sazonalidade F | Maiores altas P | Relação F e P |
| Julho a janeiro | Março e Março | ↑P ↓F e ↓P ↑F |

foco (F) e temperatura máxima (T)

|  |  |  |
| --- | --- | --- |
| Período de Sazonalidade F | Maiores altas T | Relação F e T |
| Julho a janeiro | Setembro e Outubro | ↑T ↑F e ↓T ↓F |

foco (F) e umidade relativa máxima (U)

|  |  |  |
| --- | --- | --- |
| Período de Sazonalidade F | Maiores altas U | Relação F e U |
| Julho a janeiro | Março e Março | ↑U ↓F e ↓U ↑F |

foco (F) e vento (V)

|  |  |  |
| --- | --- | --- |
| Período de Sazonalidade F | Maiores altas V | Relação F e V |
| Julho a janeiro | Setembro e Outubro | ↑V ↓F e ↓V ↑F |

Mato Grosso

foco (F) e precipitação total (P)

|  |  |  |
| --- | --- | --- |
| Período sazonalidade F | Maiores altas P | Relação F e P |
| Julho a novembro | Janeiro e fevereiro | ↑F ↓P e ↓F ↑P |

foco (F) e temperatura máxima (T)

|  |  |  |
| --- | --- | --- |
| Período sazonalidade F | Maiores altas T | Relação F e T |
| Julho a novembro | setembro | ↑T ↑F e ↓T ↓F |

foco (F) e umidade relativa máxima (U)

|  |  |  |
| --- | --- | --- |
| Período sazonalidade F | Maiores altas U | Relação F e U |
| Julho a novembro | Dezembro a fevereiro | ↑U ↓F e ↓U ↑F |

foco (F) e vento (V)

|  |  |  |
| --- | --- | --- |
| Período sazonalidade F | Maiores altas V | Relação F e V |
| Julho a novembro | agosto | Instabilidade na relação |

Pará

foco (F) e precipitação (P)

|  |  |  |
| --- | --- | --- |
| Período sazonalidade F | Maiores altas P | Relações F e P |
| Junho a janeiro | Agosto e setembro | ↑P ↓F e ↓P ↑F |

foco (F) e temperatura máxima (T)

|  |  |  |
| --- | --- | --- |
| Período sazonalidade F | Maiores altas T | Relações F e T |
| Junho a janeiro | Setembro e outubro | ↑T ↑F e ↓T ↓F |

foco (F) e umidade relativa máxima (U)

|  |  |  |
| --- | --- | --- |
| Período sazonalidade F | Maiores altas U | Relações F e U |
| Junho a janeiro | Fevereiro a abril | ↑U ↓F e ↓U ↑F |

foco (F) e velocidade (V)

|  |  |  |
| --- | --- | --- |
| Período sazonalidade F | Maiores altas V | Relações F e V |
| Junho a janeiro | Setembro e outubro | ↑V ↑F e ↓V ↓F |

Rondônia

Foco (F) e precipitação total (F)

|  |  |  |
| --- | --- | --- |
| Período de Sazonalidade F | Maiores altas P | Relações F e P |
| Junho a dezembro | Fevereiro e Fevereiro | ↑P ↓F e ↓P ↑F |

Foco (F) e temperatura máxima (T)

|  |  |  |
| --- | --- | --- |
| Período de Sazonalidade F | Maiores altas T | Relações F e T |
| Junho a dezembro | Agosto e Setembro | ↑T ↑F e ↓T ↓F |

Foco (F) e umidade relativa máxima (U)

|  |  |  |
| --- | --- | --- |
| Período de Sazonalidade F | Maiores altas U | Relações F e U |
| Junho a dezembro | Março e Dezembro | ↑U ↓F e ↓U ↑F |

Foco (F) e vento (V)

|  |  |  |
| --- | --- | --- |
| Período de Sazonalidade F | Maiores altas V | Relações F e V |
| Junho a dezembro | Agosto e Setembro | ↑V ↓F e ↓V ↑F |

Roraima

Foco (F) e precipitação total (P)

|  |  |  |
| --- | --- | --- |
| Período de Sazonalidade F | Maiores altas P | Relações F e P |
| Não definida | Junho e Junho | ↑P ↓F e ↓P ↑F |

Foco (F) e temperatura máxima (T)

|  |  |  |
| --- | --- | --- |
| Período de Sazonalidade F | Maiores altas T | Relações F e T |
| Não definida | Março e Novembro | ↑T ↑F e ↓T ↓F |

Foco (F) e umidade relativa máxima (U)

|  |  |  |
| --- | --- | --- |
| Período de Sazonalidade F | Maiores altas U | Relações F e U |
| Não definida | Julho e Junho | ↑U ↓F e ↓U ↑F |

Foco (F) e vento (V)

|  |  |  |
| --- | --- | --- |
| Período de Sazonalidade F | Maiores altas V | Relações F e V |
| Não definida | Fevereiro e Março | ↑V↓F e ↓V↑F |

Tocantins

Foco (F) e precipitação total (P)

|  |  |  |
| --- | --- | --- |
| Período de Sazonalidade F | Maiores altas P | Relações F e P |
| Julho e dezembro | Janeiro e Dezembro | ↑P↓F e ↓P↑F |

Foco (F) e temperatura máxima (T)

|  |  |  |
| --- | --- | --- |
| Período de Sazonalidade F | Maiores altas T | Relações F e T |
| Julho e dezembro | Setembro e Setembro | ↑T ↑F e ↓T↓F |

Foco (F) e umidade relativa máxima (U)

|  |  |  |
| --- | --- | --- |
| Período de Sazonalidade F | Maiores altas U | Relações F e U |
| Julho e dezembro | Março e Março | ↑U↓F e ↓U↑F |

Foco (F) e vento (V)

|  |  |  |
| --- | --- | --- |
| Período de Sazonalidade F | Maiores altas V | Relações F e V |
| Julho e dezembro | Agosto e Agosto | ↑V↓F e ↓V ↑F |

**DISCUSSÃO E CONCLUSÃO:**

Este projeto tem como finalidade buscar resolver um problema relevante da sociedade vinculada a uma das ODS ( Objetivos de Desenvolvimento Sustentável), usando Séries Temporais. Portanto, escolhemos um banco de dados do Portal TerraBrasilis sobre focos de queimadas X desmatamentos na Amazônia Legal e, também, a base de dados do INMET (Instituto Nacional de Meteorologia) que nos ajudou a desenvolver modelos de previsão, pois o nosso objetivo principal e contribuir com a diminuição desses eventos os quais causam um grande impacto sobre o clima, níveis de poluentes atmosféricos, destruição de recursos florestais, biodiversidade entre outros.

Os modelos desenvolvidos nos mostraram que as variáveis independentes, como precipitação, temperatura, umidade e velocidade do vento não influenciam diretamente nos focos de queimadas.

As últimas notícias nos revelam que há um crescimento recorde de queimadas na região estudada. Foram 8895 focos de incêndio de janeiro a abril de 2024, um crescimento de 154% em relação ao mesmo período de 2023, segundo o INPE (Instituto Nacional de Pesquisas Espaciais). Portanto, com essa análise em séries temporais, com o desenvolvimento das previsões até janeiro de 2025, podemos perceber que os focos de queimadas são influenciados por uma combinação de fatores humanos e climáticos. As atividades humanas, como desmatamento e manejo inadequado do fogo, são as principais causas iniciais dos incêndios. No entanto, as condições climáticas determinam a facilidade com que esses incêndios se espalhem e a severidade que podem causar.

Portanto a modelagem escolhida por nós, SARIMAX e VAR, nos apresentaram variações significativas no desempenho entre os estados, sendo Roraima e Tocantins os estados que se destacaram e onde os modelos apresentaram um desempenho satisfatório.

Portanto, concluímos que não há um modelo claramente superior em todos os estados. A escolha entre SARIMAX e VAR pode depender do estado especifico e das condições dos dados.

Ainda temos muitas coisas para melhorar neste projeto, como por exemplo, personalizar os modelos para cada estado, pode ser uma abordagem eficaz, considerando as dinâmicas únicas e as variáveis exógenas relevantes para cada região. Podemos, também, explorar outros modelos, assim como técnicas de Machine Learning que possam proporcionar ganhos significativos em precisão. Outro ponto importante, é continuar monitorando o desempenho dos modelos e atualizá-los regularmente com novos dados para capturar mudanças nas dinâmicas subjacentes. Aí pensamos, se não fosse melhor optarmos por fazer esse projeto em Dados IoT, pois aplicar o modelo de IoT para prever focos de queimadas e desmatamentos poderia nos oferecer várias vantagens e potencialmente melhorar a previsão dos resultados, pois a coleta dos dados em tempo real, a integração de múltiplas fontes de dados e o uso de técnicas avançadas de Machine Learning e IA poderiam oferecer insights profundos e ações práticas para enfrentar esse complexo desafio.

Logo, concluímos que o objetivo inicial referente às previsões foi atingido, mesmo com resultados não tão satisfatórios e quanto as limitações, inicialmente sentimos dificuldades de encontrar referenciais teóricos que pudessem nos auxiliar durante o projeto.

**Cronograma:**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| **PERÍODO: 06/03 a 10/03/24** | **PERÍODO: 11/03 A 07/04/24** | **PERÍODO: 08/04 A 05/05/24** | **PERÍODO: 06/05 A 30/05/24** |
|  |  |  |  |
| **ATIVIDADES:** | **ATIVIDADES:** | **ATIVIDADES:** | **ATIVIDADES:** |
|  |  |  |  |
| 1) Definição da equipe; | 1)Melhorias da etapa anterior; | 1)Melhorias das etapas anteriores; | 1)Melhorias das etapas anteriores; |
| 2) Definição da Fonte de Dados, | 2)Definida a base de dados, fazer uma | 2)Preparação dos Dados; | 2)Desenvolvimento de outro(s) |
| Banco de Dados e nome do Projeto; | breve introdução sobre o problema | 3)Análise Exploratória dos Dados; | modelo(s); |
| 3)Divisão de tarefas para cada | vinculado à ODS; | 4)Aplicação de modelo base( fase de | 3)Esclarecer limitações e dificuldades |
| integrante da equipe; | 3)Busca de Referencial Teórico | testes para verificação de resultados); | do projeto; |
| 4)Explanação sobre definição de | correlacionado ao projeto; | 5)Atualização do Diagrama de Solução; | 4)Apresentação de Conclusões |
| Séries Temporais, metodologias a | 4)Criação do Diagrama de Solução; | 6)Atualização do Cronograma; | Gerais; |
| serem utilizadas já pensando em AED; | 5)Criação do Cronograma; | 7)Atualização das Referências | 5) Apresentação doNotebook e do |
| 5)Definição de algumas Referências | 6)Entregável da Etapa 2 revisado | Bibliográficas; | Github do projeto(fase final) |
| Bibliográficas; |  | 8)Apresentação do Notebook e | 6)Vídeo de apresentação e solução |
| 6)Entregável da Etapa 1 revisado |  | Github do projeto ( fase inicial) | técnica do projeto; |
|  |  | 9)Entregável da Etapa 3 revisado | 7)Entregável da Etapa 4 revisado |

Fonte: Elaborado pelo próprio autor

**Referências Bibliográficas:**

Foram pesquisados alguns estudos, trabalhos científicos , revistas e vídeos sobre sobre Séries Temporais e a questão das queimadas e desmatamentos no Brasil , muitas explanações e algoritmos que podem auxiliar numa análise a ser realizada

FRONZA,Guilherme ; SILVA,Thaciane- Revista Galileu, publicada em 23 de junho de 2022 – “Como novas tecnologias ajudam a identificar desmatamento na Amazônia. Disponível em:<https://revistagalileu.globo.com/Ciencia/Meio-Ambiente/noticia/2022/06/como-novas-tecnologias-ajudam-identificar-desmatamento-na-amazonia.html> Acesso em:23/03/2024

CARMO,Wanderley ; CARMO, Maria das Graças – Revista Científica do Instituto Ideia – RJ , publicado em 2019 –“Desmatamento, queimadas e ameaça de extinção da flora e fauna na Amazônia Brasileira”. Disponível em: <https://revistaideario.com/pdf/desm/revista.ideario.15n.02_2019/revista.ideario.n15_02.2019.049_desmatamento.queimadas.pdf>. Acesso em :24/04/2024

CARDOSO,Jean – “Pray for Amazonia”, “Uma breve análise de Séries Temporais”, publicado em 25 de setembro de 2019. Disponível em: <https://www.linkedin.com/pulse/prayforamazonia-uma-breve-an%C3%A1lise-de-s%C3%A9ries-temporais-cardoso/> Acesso em:07/03/2024

PAPARELLI,A.Maxwell –“Time series (Séries Temporais) com Python. Disponível em: <https://maxwellpaparelli.medium.com/time-series-s%C3%A9ries-temporais-com-python-f4e74fd45b0b> Acesso em:11/03/2024

JAIN,Aarshay –“Guia completo para criar Time Series(com código em Python), publicado em 10 de outubro de 2016. Disponível em: <https://www.vooo.pro/insights/guia-completo-para-criar-time-series-com-codigo-em-python/> Acesso em 11/03/24

Vídeo aula (Parte1), “Séries Temporais com Python, Pandas ,Statsmodels” , ORMaster de 2020. Disponível em em:<https://br.video.search.yahoo.com/search/video?fr=mcafee&p=series+temporais+com+sazonalidade+e+tendencia+em+python&type=E210BR91199G91653#id=3&vid=67f27a87c200721983387b176765a8bb&action=click> Acesso em: 12/03/24

FILHO,Mario – “Machine Learning”, “Séries Temporais” , publicado em 27 de fevereiro de 2023. Disponível em: <https://mariofilho.com/tags/series-temporais/> Acesso em: 12/03/24

OLIVEIRA, R., Abarracin, OYE, Silva, GR (2024) [Introdução às Séries Temporais: Uma Abordagem Prática em Python](https://github.com/Introducao-Series-Temporais-em-Python/Book) . Editora Mackenzie.

NIELSEN,Aileen. Análise Prática de Séries Temporais. Editora Alta Books

MARTIN,Amanda Caroline;HENNING,Elisa;WALTER,M.F.C.Olga;KONRATH,Andrea Cristina- Revista Espacios, Vol.37, Número 06, pág.29,”Análise de Séries Temporais para previsão da evolução de número de automóveis no Município de Joinville”, publicado em 2016. Disponível em: <https://qualimetria.ufsc.br/files/2016/05/Revista-ESPACIOS-_-Vol.pdf> Acesso em:22/04/24

SANTOSO,B.Angga; WIDODO,Tri – Jurnal SISFOKOM, Vol 13, pag 119-129, publicado em 25 de fevereiro de 2024 –“Predicting the Number of Forest na Land Fire Hotspot Occurrences Using the ARIMA and SARIMA Methods”. Disponível em: <https://jurnal.atmaluhur.ac.id/index.php/sisfokom/article/view/2018/978> Acesso em:25/04/24

CHEN,Yang;RANDERSON,T.James “et all”, Journal of Advances in Modeling Earth Sistems – “Previsão de emissões globais de incêndios em escalas de tempo subsazonais e sazonais”, publicado em 24 de agosto de 2020. Disponível em: <https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2019MS001955> Acesso em:25/04/24

(Obs; esse link não está entrando com Ctrl+clic, tem que copiar e colar)

Sabemos que parte das queimadas e desmatamento são oriundas das ações humanas e a solução seria a conscientização e mudança de hábitos da sociedade. Mas mesmo com tantas propagandas e alertas para esse tipo de problema, as pessoas ainda cometem esse “ crime” com a natureza.

Uma solução possível seria um satélite que possa captar focos de incêndio e comunicar diretamente a central dos Bombeiros, ter leis e punições mais severas mas combinado com maior investimento em fiscalização, rejeitar produtos associados ao desmatamento, reduzir emissões globais, entre outros.