



**Universidade Presbiteriana Mackenzie**

**UNIVERSIDADE PRESBITERIANA MACKENZIE**

**TECNÓLOGO EM CIÊNCIAS DE DADOS**

Caroline Ribeiro Ferreira – 10408052

Leonardo dos Reis Olher – 10407752

Liliane Gonçalves de Brito Ferraz – 10407087

Múcio Emanuel Feitosa Ferraz Filho – 10218691

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

**SÃO PAULO**

**2024**



## Sumário

1.	GLOSSÁRIO.....	3
2.	INTRODUÇÃO .....	4
3.	MOTIVAÇÕES E JUSTIFICATIVA .....	5
4.	OBJETIVO.....	6
5.	DESCRIÇÃO DA BASE DE DADOS .....	7
6.	REFERENCIAL TEÓRICO .....	8
7.	PIPELINE DA SOLUÇÃO .....	10
8.	ANÁLISE EXPLORATÓRIA .....	12
9.	MODELO BASE.....	13
10.	CRONAGRAMA .....	15
11.	DIRETÓRIO GITHUB.....	17
7.	REFERÊNCIAS BIBLIOGRÁFICAS .....	18



## 1. GLOSSÁRIO

- **Colaboratory:** Conhecido também como “Colab”, é um produto do Google Research, área de pesquisas científicas do Google. O Colab permite que qualquer pessoa escreva e execute código Python arbitrário pelo navegador e é especialmente adequado para aprendizado de máquina, análise de dados e educação.
- **DataFrame:** É uma estrutura de dados bidimensional com os dados alinhados de forma tabular em linhas e colunas.
- **Datasets:** conjuntos de dados organizados em um formato similar ao das tabelas, com linhas e colunas que contém informações sobre determinado tema.
- **GitHub:** GitHub é uma plataforma de hospedagem de código-fonte e arquivos com controle de versão usando o Git. Ele permite que programadores, utilitários ou qualquer usuário cadastrado na plataforma contribuam em projetos privados e/ou Open Source de qualquer lugar do mundo.
- **Machine Learning (Aprendizado de máquina):** É um subcampo da Engenharia e da ciência da computação que evoluiu do estudo de reconhecimento de padrões e da teoria do aprendizado computacional em inteligência artificial.
- **Python:** É uma linguagem de programação de alto nível, interpretada de script, imperativa, orientada a objetos, funcional, de tipagem dinâmica e forte. Foi lançada por Guido van Rossum em 1991.

## 2. 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.

### **3. MOTIVAÇÕES E JUSTIFICATIVA**

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.

#### **4. 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.



## 5. 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  $-90^{\circ}$  no polo sul a  $90^{\circ}$  no polo norte, com  $0^{\circ}$  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 Pessoas – LGPD.

## 6. 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).

## 7. 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.

## 8. 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.

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

### 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;
- 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;
- Mapas, para verificar a distribuição tanto do “RiscoFogo” como do “DiaSemChuva” pelo país;
- 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, 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.

## 9. MODELO BASE

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:

- $p$  representante
- $d$  representa
- $q$  repr

A escolha dos valores de  $p, d, q$  (depois de várias tentativas).

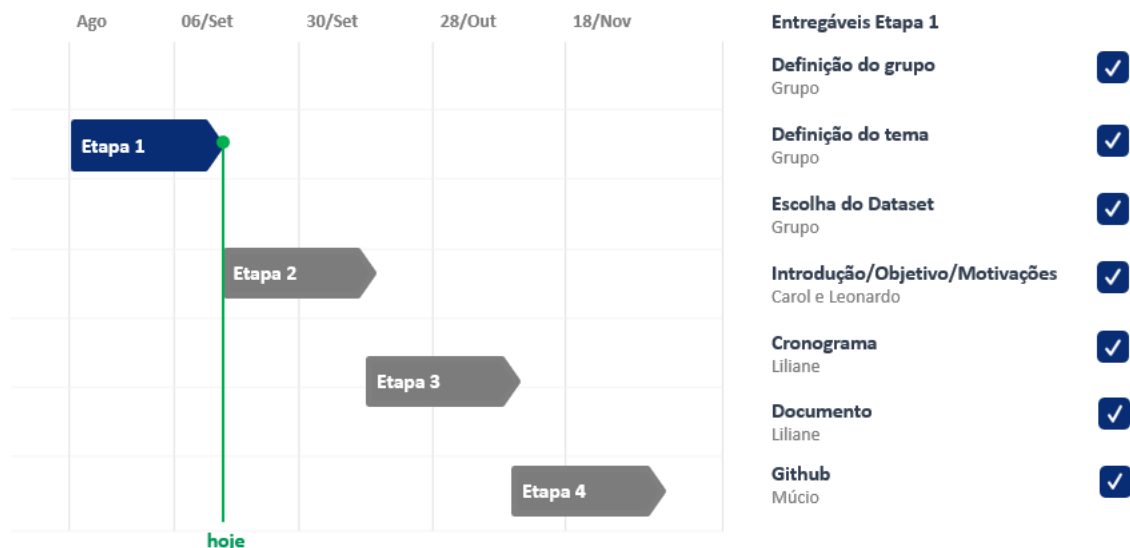
De acordo com Morettin e Toloi (2006), o modelo ARIMA é indicado para séries temporais com comportamento sazonal e tendências pronunciadas, características comuns em características naturais, como os riscos de incêndio associados a períodos de seca. Este modelo, ao capturar padrões históricos na série 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:

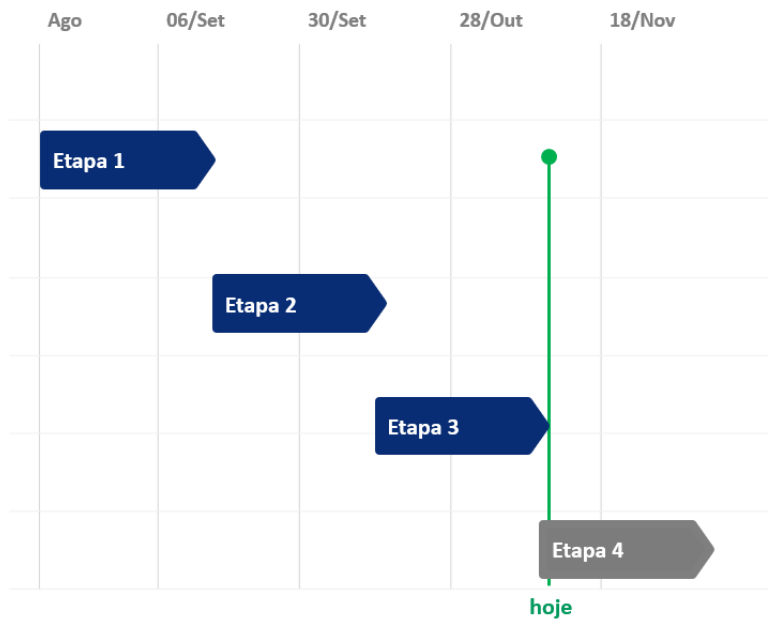
1. Seleção automática de modelos: O pmdarima pode auxiliar na escolha dos melhores hiper parâmetros para modelos ARIMA, economizando tempo e esforço.
2. Forecasting: Permite realizar previsões precisas de séries temporais, considerando a sazonalidade e tendências.
3. Análise de séries temporais: Oferece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

## 10. CRONAGRAMA

Reportar o desenvolvimento das etapas propostas e o percentual de evolução de entregas referente ao componente curricular de Projeto Aplicado VI do curso de Tecnologia em Ciências de Dados.







## Entregáveis Etapa 3

**Pesquisa sobre método ARIMA**  
Carol



**Descrição Análise Exploratória**  
Múcio



**Implementação do modelo**  
Leonardo



**Avaliação de desempenho**  
Leonardo



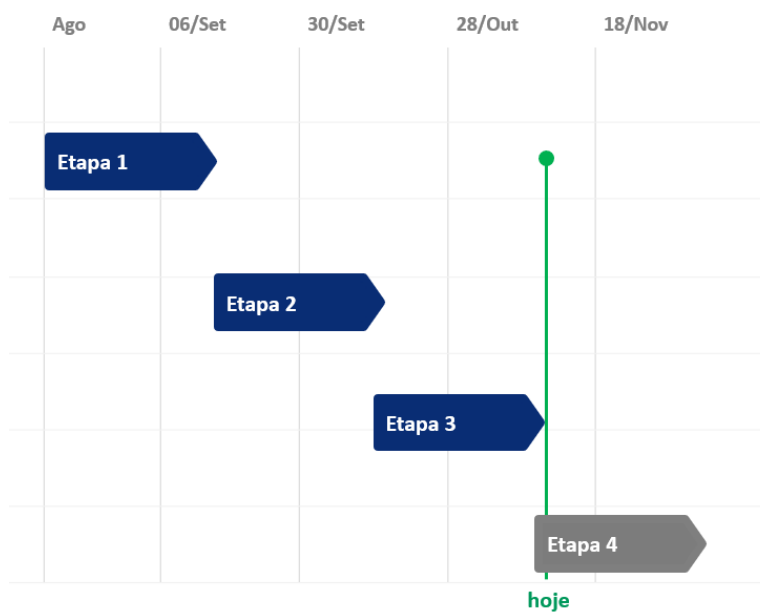
**Atualização do Cronograma**  
Liliane



**Atualização do Documento**  
Liliane



**Atualização do Github**  
Liliane



## Entregáveis Etapa 4

**Análise crítica entre os modelos**  
À definir



**Discussão dos resultados**  
À definir



**Proposição de melhorias futuras**  
À definir



**Apresentação do produto construído**  
À definir



**Atualização do Cronograma**  
À definir



**Atualização do Documento**  
À definir



**Atualização do Github**  
À definir







## 11. 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 “Códigos” será disponibilizado os códigos em Python que foram utilizados para realizar a análise exploratória, tratamento dos dados e o estudo de series temporais.

Pasta “Dados” temos os arquivos utilizados para o estudo.

Pasta “Documentos” temos o cronograma de entrega do projeto e as versões de entrega deste documento.

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

## 7. 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.



➤ Sites de consultas:

<https://www.ipea.gov.br/ods/ods13.html>

<https://terrabrasilis.dpi.inpe.br/queimadas/bdqueimadas/#exportar-dados>

<https://analisemacro.com.br/data-science/analise-do-foco-de-queimadas-no-brasil-usando-python/>

<https://www.poder360.com.br/meio-ambiente/area-queimada-no-brasil-cresce-410-em-fevereiro-diz-estudo/>

<https://ipam.org.br/fogo-queimou-quase-200-milhoes-de-hectares-do-brasil-nos-ultimos-39-anos/>

<https://agenciabrasil.ebc.com.br/geral/noticia/2024-02/area-queimada-no-brasil-cresce-248-em-relacao-janeiro-de-2023>

<https://www.neoenergia.com/w/queimadas-no-brasil-por-que-devemos-nos-preocupar-#:~:text=Para%20controlar%20um%20inc%C3%AAndio%2C%20duas,nas%20%C3%A1reas%20afetadas%20pelo%20inc%C3%AAndio>

[\[coids.inpe.br/queimadas/queimadas/riscofogo\\\_meteorologia/anuario\\\_risco\\\_de\\\_fogo/anuario\\\_risco\\\_de\\\_fogo\\\_2023.pdf\]\(https://coids.inpe.br/queimadas/queimadas/riscofogo\_meteorologia/anuario\_risco\_de\_fogo/anuario\_risco\_de\_fogo\_2023.pdf\)](https://dataserver-</a></p></div><div data-bbox=)

<https://terrabrasilis.dpi.inpe.br/queimadas/porta/faq/index.html#:~:text=7.,regi%C3%B5es%20e%20per%C3%ADodos%20de%20interesse>

<https://educa.ibge.gov.br/jovens/conheca-o-brasil/territorio/18307-biomas-brasileiros.html>

<https://www.terra.com.br/planeta/noticias/media-de-dias-seguidos-sem-chuva-no-brasil-aumentou-de-80-para-100-em-60-anos-aponta->

[inpe.1c8479af867aef5dc423c84a9bc26900m1rhaj8p.html#:~:text=O%20n%C3%BAmero%20de%20dias%20consecutivos,n%C3%BAmeros%20de%201960%20a%202020](https://inpe.1c8479af867aef5dc423c84a9bc26900m1rhaj8p.html#:~:text=O%20n%C3%BAmero%20de%20dias%20consecutivos,n%C3%BAmeros%20de%201960%20a%202020)

[https://pt.wikipedia.org/wiki/Precipita%C3%A7%C3%A3o\\_\(meteorologia\)](https://pt.wikipedia.org/wiki/Precipita%C3%A7%C3%A3o_(meteorologia))

<https://pt.wikipedia.org/wiki/Latitude>

<https://pt.wikipedia.org/wiki/Longitude>

<https://www.nationalgeographicbrasil.com/meio-ambiente/2024/09/como-as-queimadas-afetam-o-ciclo-da-agua-o-resultado-pode-ser-uma-cri-se-de-abastecimento>

<https://mtc-m21c.sid.inpe.br/col/sid.inpe.br/mtc-m21c/2019/11.21.11.03/doc/publicacao.pdf>