

UNIVERSIDADE PRESBITERIANA MACKENZIE
FACULDADE DE COMPUTAÇÃO E INFORMÁTICA

Modelagem Preditiva de Incêndios Florestais no Brasil

São Paulo
2024

Sumário

1. TÍTULO ESCOLHIDO:	3
2. MEMBROS:	3
3. CONTEXTO DO ESTUDO:	4
3.1 O Brasil como Foco:	4
3.2 Instituição Escolhida: Instituto Nacional de Pesquisas Espaciais (INPE)	5
4. PREMISSAS DO PROJETO:	6
5. OBJETIVOS E METAS	7
6. METODOLOGIA	7
7. CRONOGRAMA DE ATIVIDADES	9
8. DEFINIÇÃO DO PRODUTO ANALÍTICO	9
8.1 Descrição e Entregáveis do Produto Analítico: O produto analítico sob desenvolvimento é um modelo preditivo de incêndios florestais baseado em dados históricos e informações ambientais, através de técnicas de aprendizado de máquina.	9
8.2 Definição de Bibliotecas (Pacotes) da Linguagem Python e Implementação no GitHub: Para dar início aos trabalhos colaborativos, definimos as bibliotecas Python a serem utilizadas e configuramos um repositório no GitHub. As bibliotecas utilizadas são:	9
8.3 Coleta e Preparação de Dados:	10
8.4 Análise Exploratória de Dados: Realizamos uma análise exploratória dos dados para compreender suas características e padrões.	10
8.5 Modelagem Preditiva:	11
8.6 Explicação do Código em desenvolvimento: O código é uma parte essencial do desenvolvimento do produto analítico. E está estruturado da seguinte forma:	12
9. REPOSITÓRIO	13
10. REFERÊNCIAS BIBLIOGRÁFICAS	13

1. TÍTULO ESCOLHIDO:

Modelagem Preditiva de Incêndios Florestais no Brasil.

2. MEMBROS:

Nome: Camila Vieira

RA: 10414794

E-mail: 10414794@mackenzista.com.br

Nome: Gabriel Schonenberger de Campos

RA: 10415150

E-mail: 10415150@mackenzista.com.br

Nome: Glayton de Paula

RA: 10415099

E-mail: 10415099@mackenzista.com.br

Nome: João Victor Mendes Cunha

RA: 10415459

E-mail: 10415459@mackenzista.com.br

3. CONTEXTO DO ESTUDO:

3.1 O Brasil como Foco:



O Brasil, com sua vasta extensão territorial e diversidade ambiental, enfrenta desafios significativos relacionados à ocorrência de incêndios florestais. Em 2021, o país foi particularmente afetado por um aumento alarmante no número de queimadas, especialmente na região da Amazônia, que registrou um aumento de 16% no número de focos de incêndio em comparação com o ano anterior, segundo dados do Instituto Nacional de Pesquisas Espaciais (INPE, 2021).

Esses eventos promovem impactos devastadores no meio ambiente, na biodiversidade e consequentemente, na sociedade como um todo. Segundo De Hoyos Guevara (2019), as florestas tropicais, como a Amazônia, são importantes para o mundo inteiro porque capturam e armazenam o carbono da atmosfera. A Amazônia também, exerce um papel fundamental no clima da América do Sul por seu efeito sobre o ciclo hidrológico local. No entanto, quando ocorrem incêndios nessas áreas, grandes quantidades de carbono são liberadas na atmosfera, contribuindo significativamente para o aquecimento global e mudanças climáticas.

Além disso, os incêndios florestais promovem impactos diretos na biodiversidade, devastando habitats naturais e colocando em risco a sobrevivência de espécies. Muitas vezes, as áreas afetadas pelos incêndios são de difícil acesso, o que dificulta o trabalho de resgate de animais e controle do fogo. Como mencionado por Silva et al. (2020), a biodiversidade da Amazônia está entre as mais ricas do mundo, e cada incêndio representa uma ameaça significativa a essa diversidade única.

Do ponto de vista socioeconômico, os incêndios florestais podem causar prejuízos significativos às comunidades locais que dependem dos recursos naturais das florestas para sua subsistência. Além disso, como abordado por Silva et al. (2020), a fumaça produzida pelos incêndios aumenta o risco de doenças respiratórias e problemas de saúde, sobrecarregando os sistemas de saúde locais e prejudicando ainda mais as comunidades afetadas.

Diante desse cenário, torna-se essencial desenvolver estratégias eficazes de prevenção e combate a incêndios. A utilização de técnicas como aprendizado de máquina e análise estatística, pode contribuir significativamente para a identificação de padrões e tendências nos dados de incêndios florestais, permitindo a adoção de medidas preventivas e a tomada de decisões assertivas (baseadas em dados) na gestão desses eventos.

3.2 Instituição Escolhida: Instituto Nacional de Pesquisas Espaciais (INPE)



Missão: Contribuir para o conhecimento e o desenvolvimento científico e tecnológico do país nas áreas espacial e ambiental.

Visão: Ser uma referência nacional e internacional em pesquisas espaciais e ambientais, promovendo a sustentabilidade e a preservação do meio ambiente.

Segmento de Atuação do INPE: O INPE é responsável pela geração, processamento e disseminação de informações e imagens de sensoriamento remoto, incluindo dados sobre queimadas e incêndios florestais. A instituição desempenha um papel fundamental no monitoramento e na análise desses eventos, fornecendo subsídios para políticas públicas e ações de conservação ambiental.

Iniciativas na Área de Ciência de Dados: O INPE tem investido em iniciativas de ciência de dados aplicadas à análise de imagens de satélite e dados ambientais. A utilização de técnicas avançadas de processamento de imagens e aprendizado de máquina tem possibilitado a identificação e a previsão de padrões de incêndios florestais com maior precisão e rapidez.

Trabalhos em Destaque: O INPE realiza pesquisas e desenvolve ferramentas para monitoramento e modelagem de incêndios florestais, como o Sistema de Detecção de Desmatamento em Tempo Real (DETER) e o Projeto de Monitoramento do Desmatamento na Amazônia Legal por Satélite (PRODES). Essas iniciativas têm contribuído significativamente para a compreensão e o controle da devastação ambiental causada por incêndios e desmatamentos.

Ao utilizar os recursos e dados disponibilizados pelo INPE, este projeto visa desenvolver um modelo preditivo de incêndios florestais baseado em dados históricos e em informações ambientais, contribuindo para a prevenção e mitigação desses eventos no Brasil.

4. PREMISSAS DO PROJETO:

- Utilização de dados fornecidos pelo INPE como fonte confiável e abrangente de informações sobre incêndios florestais.

- Utilizar técnicas de manipulação de texto.
- Associar o estudo as demais disciplinas do terceiro semestre do curso de Ciência de Dados, como estatística e aprendizado de máquina, para análise e modelagem preditiva de incêndios florestais.
- Consideração de variáveis ambientais, climáticas e geoespaciais para uma análise abrangente e precisa.
- Foco na previsão de padrões de incêndios florestais com base em dados históricos e em informações em tempo real.
- Fornecer relatórios que sirvam como apoio em possível tomada de decisão.

5. OBJETIVOS E METAS

Objetivo Principal: Desenvolver um modelo preditivo de incêndios florestais utilizando técnicas de manipulação de dados textuais.

Metas:

- Coletar e processar dados históricos e em tempo real sobre incêndios florestais.
- Identificar padrões e tendências nos dados relacionados à ocorrência de incêndios florestais.
- Desenvolver e validar modelos de previsão de incêndios florestais com base em variáveis ambientais e climáticas.
- Avaliar a eficácia do modelo preditivo na identificação e previsão de incêndios florestais em diferentes regiões do Brasil.

6. METODOLOGIA

Neste projeto, adotou-se uma abordagem baseada no pensamento computacional. A Computing at School (2015) relaciona ao PC as habilidades de raciocínio lógico, pensamento algorítmico, decomposição, generalização, reconhecimento de padrões, abstração, representação e avaliação e presente em todas as áreas, o Pensamento Computacional está intrinsecamente ligado à resolução de problemas” (SBC, 2019, p.5).

As características do pensar computacional privilegiam elementos do saber e do fazer matematicamente, como: formular problemas; representar dados através de abstrações, como modelos e simulações; automatizar soluções através do pensamento algorítmico; identificar, analisar e implementar possíveis soluções; lidar com problemas abertos e imprevisíveis, como: abstração, algoritmo, decomposição, reconhecimento e generalizações de padrões, etc. (BARBA, 2016; WING, 2014).

A partir disso, estruturamos o projeto da seguinte forma:

Decomposição: Iniciaremos decompondo o problema complexo de previsão de incêndios florestais em partes menores e gerenciáveis. Dividiremos o processo em etapas distintas, como coleta de dados, pré-processamento, modelagem e avaliação, para facilitar o entendimento e a resolução do problema.

Reconhecimento de Padrões: Utilizaremos técnicas de reconhecimento de padrões para identificar regularidades nos dados relacionados a incêndios florestais. Por meio de análises exploratórias e visualizações, procuraremos padrões significativos que possam influenciar a ocorrência desses eventos, como sazonalidade, características geográficas e fatores climáticos.

Abstração: Com base nos padrões identificados, realizaremos uma abstração dos dados, simplificando aspectos complexos para focar nos elementos essenciais relacionados à ocorrência de incêndios florestais. Agruparemos variáveis relevantes e criaremos representações simplificadas que capturam a essência do fenômeno, facilitando a análise e o desenvolvimento do modelo preditivo.

Algoritmo: Implementaremos algoritmos de aprendizado de máquina e técnicas estatísticas para desenvolver o modelo preditivo de incêndios florestais. Utilizaremos algoritmos de classificação, regressão e redes neurais, para treinar o modelo com os dados históricos e prever a ocorrência de incêndios em novas situações.

7. CRONOGRAMA DE ATIVIDADES

AÇÃO	FASE	COMPLEXIDADE	EXPECTATIVA	TEMPO ESTIMADO	STATUS
Criação do grupo	FASE 1	BAIXA	OTIMISTA	1,333333333	CONCLUÍDO
Definição da premissas	FASE 1	BAIXA	MODERADA	2	CONCLUÍDO
Definição de objetivos e metas	FASE 1	BAIXA	MODERADA	2	CONCLUÍDO
Definição do cronograma de atividades	FASE 1	BAIXA	MODERADA	2	CONCLUÍDO
Organização do documento da etapa 1	FASE 1	BAIXA	OTIMISTA	1,333333333	CONCLUÍDO
Definição de bibliotecas a serem usadas	FASE 2	MÉDIA	MODERADA	6	CONCLUÍDO
tratamento da base de dados	FASE 2	ALTA	MODERADA	12	CONCLUÍDO
Definição e descrição da base teórica	FASE 2	MÉDIA	MODERADA	6	CONCLUÍDO
Definição e descrição do cálculo de acurácia	FASE 2	MÉDIA	MODERADA	6	CONCLUÍDO
Organização do documento da etapa 2	FASE 2	BAIXA	OTIMISTA	1,333333333	CONCLUÍDO
Aplicação do método analítico escolhido	FASE 3	MÉDIA	MODERADA	6	A SER INICIADO
Cálculo das medidas de acurácia	FASE 3	ALTA	MODERADA	12	A SER INICIADO
Descrição dos resultados preliminares apresentando produto gerado	FASE 3	MÉDIA	MODERADA	6	A SER INICIADO
Esboço do <i>storytelling</i>	FASE 3	MÉDIA	MODERADA	6	A SER INICIADO
Organização do documento da etapa 3	FASE 3	BAIXA	OTIMISTA	1,333333333	A SER INICIADO
Finalização do relatório técnico	FASE 4	MÉDIA	MODERADA	6	A SER INICIADO
Criar apresentação do <i>storytelling</i>	FASE 4	MÉDIA	OTIMISTA	4	A SER INICIADO
Disponibilização do repositório GitHub do projeto	FASE 4	BAIXA	OTIMISTA	1,333333333	A SER INICIADO
Gravação e disponibilização do vídeo de apresentação final do projeto	FASE 4	MÉDIA	MODERADA	6	A SER INICIADO

8. DEFINIÇÃO DO PRODUTO ANALÍTICO

Nesta fase do projeto, definimos o produto analítico que será desenvolvido com base na análise e modelagem dos dados sobre incêndios florestais.

8.1 Descrição e Entregáveis do Produto Analítico: O produto analítico sob desenvolvimento é um modelo preditivo de incêndios florestais baseado em dados históricos e informações ambientais, através de técnicas de aprendizado de máquina.

8.2 Definição de Bibliotecas (Pacotes) da Linguagem Python e Implementação no GitHub: Para dar início aos trabalhos colaborativos, definimos as bibliotecas Python a serem utilizadas e configuramos um repositório no GitHub. As bibliotecas utilizadas são:

NumPy e Pandas: essenciais para manipulação e análise de dados.

Matplotlib e Seaborn: utilizadas para visualização de dados.

Scikit-learn: importante para implementação de algoritmos de aprendizado de máquina.

GeoPandas e PyOGGIO: utilizada para manipulação e análise de dados geoespaciais.

8.3 Coleta e Preparação de Dados: Utilizamos dados fornecidos pelo INPE, incluindo informações sobre risco de fogo, precipitação e localização geográfica dos focos de incêndio.

Realizamos a leitura e preparação dos dados utilizando as bibliotecas geopandas e pandas.

```
In [ ]: !pip install geopandas
        !pip install pyogrio

In [100... # prompt: importar biblioteca geopandas
from osgeo import gdal
import geopandas as gpd
import pandas as pd
import pyogrio;
```

8.4 Análise Exploratória de Dados: Realizamos uma análise exploratória dos dados para compreender suas características e padrões.

```
In [70]: qsp = gpd.read_file('focos_qmd_inpe_2024-03-22_2024-03-23.shp', engine='pyogrio')
```

```
In [71]: qsp.head()
```

```
Out[71]: geometry
0 POINT (-61.11770 3.51040)
1 POINT (-61.09910 3.49220)
2 POINT (-61.71040 3.05250)
3 POINT (-61.59880 2.96140)
4 POINT (-60.27440 2.92790)
```

```
In [94]: qsp['geometry'].size
```

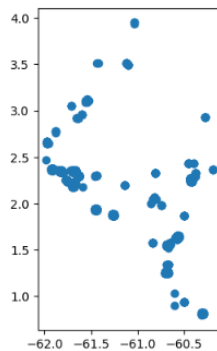
```
Out[94]: 784
```

```
In [89]: qsp.dtypes
```

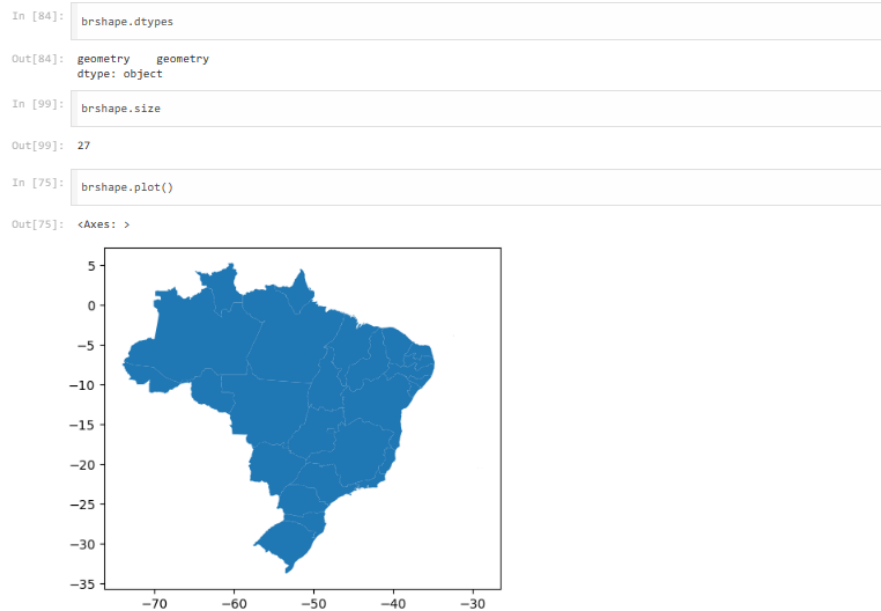
```
Out[89]: geometry geometry
dtype: object
```

```
In [73]: qsp.plot()
```

```
Out[73]: <Axes: >
```



Visualizamos os dados geoespaciais em mapas para identificar áreas com maior incidência de incêndios e examinamos estatísticas descritivas e gráficos para entender a distribuição e correlação entre as variáveis.



8.5 Modelagem Preditiva:

- Utilizamos o algoritmo RandomForestRegressor do Scikit-learn para construir o modelo preditivo.
- Dividimos os dados em conjuntos de treinamento e teste.
- Ajustamos o modelo aos dados de treinamento e avaliamos seu desempenho usando métricas como o erro médio quadrático.

```

In [102]: dfq = pd.read_csv('focos_qmd_inpe_2024-03-22_2024-03-23.csv')

In [78]: from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error

In [104]: dfq = dfq[dfq['RiscoFogo'].notnull() & dfq['Precipitacao'].notnull() & dfq['DiaSemChuva'].notnull()]

In [105]: dfq.head()

Out[105]:
   DataHora  Satelite Pais Estado Municipio Bioma DiaSemChuva Precipitacao RiscoFogo Latitude Longitude FRP
8189 2024/03/22 00:04:43 GOES-16 Brasil RORAIMA AMAJARI Amazônia 5.0 0.0 0.98 3.6916 -61.6808 54.7
8190 2024/03/22 00:04:46 GOES-16 Brasil RORAIMA ALTO ALEGRE Amazônia 5.0 0.0 1.00 3.1439 -61.6148 58.1
8191 2024/03/22 00:04:46 GOES-16 Brasil RORAIMA ALTO ALEGRE Amazônia 5.0 0.0 1.00 3.1257 -61.6151 124.7
8192 2024/03/22 00:04:46 GOES-16 Brasil RORAIMA ALTO ALEGRE Amazônia 5.0 0.0 1.00 3.1257 -61.5962 75.3
8193 2024/03/22 00:04:46 GOES-16 Brasil RORAIMA ALTO ALEGRE Amazônia 15.0 0.0 1.00 2.8150 -61.8080 64.3

In [106]: dfq.size

Out[106]: 132900

In [111]: X = dfq[['DiaSemChuva', 'Precipitacao', 'Latitude', 'Longitude']]
y = dfq['RiscoFogo']

In [112]: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

In [113]: model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

Out[113]: RandomForestRegressor(random_state=42)

```

8.6 Explicação do Código em desenvolvimento: O código é uma parte essencial do desenvolvimento do produto analítico. E está estruturado da seguinte forma:

1. Importação de bibliotecas e leitura de dados geoespaciais: as primeiras linhas do nosso código estão importando as bibliotecas necessárias, como geopandas, e lendo arquivos de dados geoespaciais, o shapefile `focos_qmd_inpe_2024-03-22_2024-03-23.shp`, que contém informações sobre focos de incêndio, retirado da base de dados do INPE.
2. Exploração inicial dos dados geoespaciais: após a leitura do arquivo shapefile, o código imprime um resumo dos dados, como as primeiras linhas (`qsp.head()`), as dimensões (`qsp['geometry'].size`) e os tipos de dados (`qsp.dtypes`). Além disso, também imprime os dados (`qsp.plot()`) que mostram um mapa com os pontos de focos de incêndio.
3. Leitura e preparação de dados: ao ler os dados (`focos_qmd_inpe_2024-03-22_2024-03-23.csv`) sobre incêndios florestais, o código os filtra para remover possíveis linhas com valores nulos.
4. Modelagem: utilizamos o algoritmo `RandomForestRegressor` do Scikit-learn para modelagem preditiva. Ele divide os dados em conjuntos de treinamento e teste, ajusta o modelo aos dados de treinamento e, em seguida, calcula a métrica de erro médio quadrático (`mean_squared_error`) para avaliar o desempenho do modelo.

5. Testagem e cálculo de acurácia: utilizamos as ferramentas MSE e r^2_score para testar a acurácia do modelo treinado. O MSE indica a média dos quadrados dos erros entre as previsões e os valores reais, enquanto o R^2 representa a proporção da variância para a variável dependente que é explicada por suas variáveis independentes no modelo.

9. REPOSITÓRIO

<https://github.com/glaytonpauladev/firerobot>

10. REFERÊNCIAS BIBLIOGRÁFICAS

BARBA, L. A. *Computacional thinking*: I do not think it means what you think it means. [2016]. Disponível em: <https://cutt.ly/lqgjorj> Acesso em: 25 de março de 2024.

Computing at School, a subdivision of the British Computer Society (BCS). (2015). *Computational Thinking: A Guide for Teachers*. Disponível em: <http://community.computingatschool.org.uk/files/6695/original.pdf>. Acesso em: 25 de março de 2024.

DE HOYOS GUEVARA, Arnoldo José et al. AÇÕES CONTRA A MUDANÇA GLOBAL DO CLIMA. Disponível em: <https://www.pucsp.br/sites/default/files/download/eventos/bisus/7-mudanca-global-climatica.pdf>. Acesso em: 10 de março de 2024.

SBC. Diretrizes para ensino de Computação na Educação Básica. 2019. Disponível em: <https://www.sbc.org.br/educacao/diretrizes-para-ensino-de-computacao-na-educacao-basica>. Acesso em: 25 de março de 2024.

Silva, A., Oliveira, B., Santos, C., & Souza, D. (2020). Biodiversidade da Amazônia: importância e desafios. *Revista Brasileira de Biociências*, 18(2), 224-231. Disponível em: <https://www.scielo.br/j/jbpneu/a/ymg4hVTN8ftSgwKVL3YC7Zt/?format=pdf&lang=pt>. Acesso em: 10 de março de 2024.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE). Monitoramento de Queimadas e Incêndios. Disponível em: <https://queimadas.dgi.inpe.br/queimadas/>. Acesso em: 25 de fevereiro de 2024.

SILVA, L. V. N. Promoção de bioeconomia da sociobiodiversidade amazônica: o caso da Natura Cosméticos S/A com comunidades agroextrativistas na região do Baixo Tocantins no Pará. 2020. Dissertação (Mestrado profissional MPGC) – Escola de Administração de Empresas de São Paulo, Fundação Getúlio Vargas, São Paulo, 2020.

Disponível

em:

https://www.google.com/search?q=%3Chttps%3A%2F%2Fbibliotecadigital.fgv.br%2Fdspace%2Fbitstream%2Fhandle%2F10438%2F29974%2FSilva%252c%2520Luciana%2520Villa%2520Nova%2520v%2520final.pdf%3Fsequence%3D1%26isAllowed%3Dy%3E&rlz=1C1GCEU_pt-BRBR1059BR1059&oq=%3Chttps%3A%2F%2Fbibliotecadigital.fgv.br%2Fdspace%2Fbitstream%2Fhandle%2F10438%2F29974%2FSilva%252c%2520Luciana%2520Villa%2520Nova%2520v%2520final.pdf%3Fsequence%3D1%26isAllowed%3Dy%3E&gs_lcrp=EgZjaHJvbWUyBggAEEUYOdIBBzU4MmowajSoAgCwAgA&sourceid=chrome&ie=UTF-8. Acesso em: 10 de março de 2024.

Silva, A., Oliveira, B., Santos, C., & Souza, D. (2020). Biodiversidade da Amazônia: importância e desafios. *Revista Brasileira de Biociências*, 18(2), 224-231. Disponível em: <https://www.scielo.br/j/jbpneu/a/ymg4hVTN8ftSgwKVL3YC7Zt/?format=pdf&lang=pt>. Acesso em: 10 de março de 2024.