

Queimadas no Pantanal brasileiro: uma análise exploratória da última década

Charles Pimentel¹, Isaac D'Césares¹

¹ Programa de Pós-Graduação em Informática
Universidade Federal do Rio de Janeiro (UFRJ)

pimentelufrj@gmail.com, isaac.dcesares@ufrj.br

Abstract. *This exploratory research aims to provide a reflection on the fires in the Pantanal biome in the last 10 years, as a means of collaborating to raise society's awareness of the urgency of preserving the region, in addition to fighting and preventing future fire outbreaks. Through a partnership with researchers from LASA – Laboratory of Environmental Satellite Applications – UFRJ, the data obtained, structured, and analyzed pointed to 2020 as a year in which fires occurred with greater intensity in the region. During the research, comparative analyzes were carried out over the last 10 years, and, through a machine learning algorithm, predictive analyses for 2022 were performed. In addition to the construction of a dynamic dashboard, to facilitate the visualization of the non-technical public, the work presents an important contribution by converting the data originally obtained in .shp format, to the .csv format, widely used in the field of Data Science.*

Resumo. *Esta pesquisa exploratória tem como objetivo proporcionar uma reflexão sobre as queimadas no bioma Pantanal nos últimos 10 anos, como meio de colaborar para conscientização da sociedade quanto a urgência de se preservar a região, além de combater e prevenir futuros focos de incêndio. Por meio da parceria com pesquisadores do LASA – Laboratório de Aplicações de Satélites Ambientais – UFRJ, os dados obtidos, estruturados e analisados apontaram para 2020 como um ano onde os incêndios ocorreram com maior intensidade na região. Durante a pesquisa, foram realizadas análises comparativas ao longo dos últimos 10 anos e, por meio de um algoritmo de aprendizado de máquina, uma análise preditiva para 2022 foi realizada. Além da construção de um dashboard dinâmico, para facilitar a visualização do público não técnico, o trabalho apresenta uma importante contribuição ao converter os dados originalmente obtidos no formato .shp, para o formato .csv, muito utilizado na área de Ciência de Dados.*

1. Introdução

No Brasil, ao longo da última década, as queimadas ocorridas nos seus variados biomas são motivo de preocupação de entidades e cidadãos que atuam a favor da preservação do Meio Ambiente, tanto no Brasil, quanto no exterior.

Um dos biomas cujo as queimadas têm sido recentemente amplamente noticiadas pela mídia nacional e internacional, é o Pantanal. As intensas queimadas ocorridas no ano de 2020 foram destacadas em inúmeras publicações, que alertaram a população em

relação ao número de queimadas que ocorreram em anos anteriores. Segundo dados publicados no site brasileiro da World Wildlife Fund¹, entre 01 de janeiro e 12 de julho de 2020 foram detectados 2510 focos de incêndio, um número 126% maior que o mesmo período em 2019.

As queimadas podem ser desencadeadas por ações climáticas ou antropogênicas, e a perda no conjunto de vida vegetal e animal é irreparável [Li et al. 2020]. A partir dessa questão, emerge a necessidade de ações de prevenção para que atitudes possam ser tomadas por organizações governamentais e não governamentais, com o objetivo de proteger as riquezas naturais desse bioma. Para esse trabalho foi utilizado o dataset disponibilizado pelo Laboratório de Aplicações de Satélites Ambientais (LASA) do Departamento de Meteorologia da UFRJ.

Desta forma, este artigo procura contribuir com este debate, apresentando um trabalho em desenvolvimento numa disciplina do Programa de Pós-Graduação em Informática na Universidade Federal do Rio de Janeiro, onde se utiliza recursos de Ciências de Dados em aplicações que envolvem problemas da sociedade atual.

Assim, na seção 2, apresentamos Trabalhos relacionados, na seção 3 os Materiais e métodos e, na seção 4, destacamos como conduzimos o Experimento. A seção 5 discutimos os princípios FAIR neste trabalho, na seção 6 apresentamos os Resultados e, por fim, na seção 7 tratamos das Considerações finais e dos trabalhos futuros.

2. Trabalhos relacionados

O tema relacionado a queimadas no bioma Pantanal tem sido explorado por pesquisadores que têm tornado público os impactos da destruição da sua fauna e da sua flora, causados tanto por aspectos climáticos, quanto por aspectos antropogênicos. Essas pesquisas contribuem para que o governo e a sociedade civil brasileira possam tomar iniciativas para impedir o avanço da destruição deste importante patrimônio natural, um dos mais ricos em biodiversidade do planeta, e apontam para a importância da preservação do ecossistema local.

[Leal Filho et al. 2021] realçam que no ano de 2020 foram registrados 15000 focos de incêndio na região do Pantanal, contra 4940 focos em 2019, apontando o mês de setembro deste mesmo ano, que registrou um aumento de 215% em relação ao mesmo período do ano anterior.

As queimadas ocorridas em 2020 não tiveram precedentes, como afirma [Marques et al. 2021], em seu trabalho publicado no Journal of Environmental Management. Os autores apontam para como as atividades antrópicas que impactam o clima e uso e cobertura de solo afetam a dinâmica espacial, temporal e de magnitude dos focos de incêndio, ocasionando, por exemplo, uma tendência decrescente para a humidade do solo.

[Correa et al. 2022] mostram que os dados relacionados a incêndios na região pantaneira, coletados nas últimas duas décadas, destacam que as queimadas costumam ocorrer mais frequentemente em pastagens do que em outros tipos de cobertura de solo, mas que em 2020 as queimadas ocorreram com mais intensidade as regiões florestais.

¹https://www.wwf.org.br/natureza_brasileira/areas_prioritarias/pantanal/pantanal_news/76708/Pantanal-tem-126-mais-queimadas-que-em-2019

No artigo denominado *Rescue Brazil's burning Pantanal wetlands* (Resgatar o Pantanal em chamas no Brasil), publicado em 2020 na importante Revista *Nature*, os autores destacaram que em 2020 um terço da região do Pantanal, o maior pântano tropical do mundo, teve um terço de sua região incendiada [Libonati et al. 2020]. Além disso, o trabalho aponta a necessidade de que medidas sejam tomadas pelo governo brasileiro para conter os incêndios nessa região, tais como gestão eficaz do fogo e políticas de proteção ambiental.

Estudos e publicações sobre os incêndios ocorridos no ano de 2020 tem um importante destaque nos trabalhos supracitados, pois foi um ano que serve de alerta para que políticas públicas de preservação, prevenção e combate, possam ser implementadas. Além disso, para que esses estudos e reflexões sejam possíveis, os dados relacionados a focos de incêndio, coletados e disponibilizados por instituições como o LASA, são de vital importância para essa finalidade.

3. Materiais e métodos

3.1. Definição do problema e objetivos

Para orientar a solução proposta, foi utilizada a metodologia CRISP-DM, a qual é um modelo de processo que fornece uma estrutura completa e iterativa para projetos de mineração/ciência de dados [Wirth and Hipp 2000].

A CRISP-DM, acrônimo de Cross Industry Standard Process for Data Mining, foi desenvolvida por uma equipe de especialistas em mineração de dados da indústria, academia e governo que buscava definir um padrão de processo de mineração de dados que pudesse ser utilizado por qualquer organização. O framework é composto por seis fases, sendo elas: compreensão do negócio, compreensão dos dados, preparação dos dados, modelagem, avaliação e implantação. A Figura 1 apresenta o modelo de processo CRISP-DM.

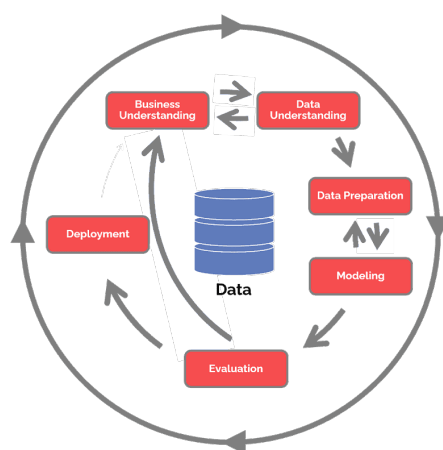


Figura 1. Representação gráfica da metodologia CRISP-DM.

Como primeira etapa do fluxo de trabalho de um projeto de Ciência de Dados, conforme a fundamentação metodológica mencionada, a definição do problema a ser resolvido e os objetivos a serem alcançados é fundamental. Através disso, é possível definir

o escopo do projeto, ou seja, o que será e o que não será feito e o que é necessário para que o projeto seja bem-sucedido.

Este trabalho deseja responder a seguinte pergunta norteadora: É possível prever as queimadas de áreas e suas extensões através da análise de série histórica das áreas queimadas na região do Pantanal, por meio de padrões encontrados por um algoritmo de aprendizado de máquina, embora as queimadas, para ocorrerem, tenham tanto o viés natural quanto o viés antropogênico?

Através da formalização do problema foi possível definir os objetivos do projeto, organizados em três níveis de prioridade que serão abordados na seção seguinte.

3.2. Definição do problema e objetivos

A solução proposta para o problema apresentado é composta por três etapas principais, conforme ilustrado na Figura 2. A primeira etapa consiste em realizar uma análise exploratória da série histórica, a fim de encontrar padrões que possam ser utilizados para prever as queimadas de áreas e suas extensões. A segunda etapa apresenta a utilização de um algoritmo de aprendizado de máquina para prever as queimadas de áreas, a partir dos padrões encontrados na análise da série histórica. A terceira etapa consolida uma comparação dos resultados obtidos com os dados reais de queimadas de áreas, a fim de verificar a eficácia do algoritmo.

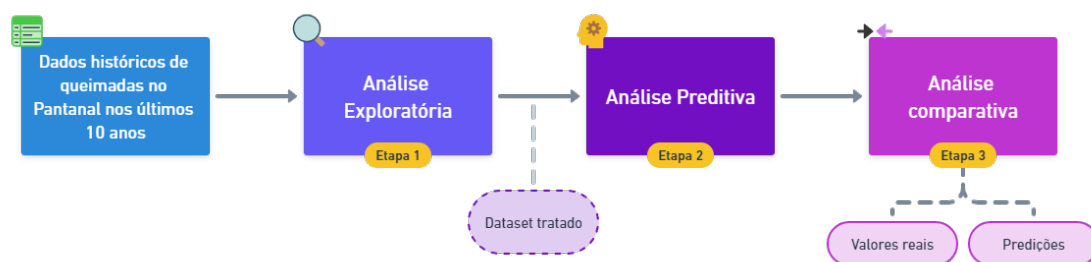


Figura 2. Representação gráfica da metodologia CRISP-DM.

4. Experimento

4.1. Proveniência

Em um projeto de ciência de dados é importante relatar a proveniência dos dados, ou seja, origem, mudanças e detalhes que suportem a verificação de sua qualidade e a confiabilidade dos resultados obtidos. Para isso, existem padrões genéricos de modelos de dados que descrevem minuciosamente informações de proveniência a fim de apoiar o intercâmbio interoperacional de informação em ambientes heterogêneos como a Internet.

Neste projeto, todos os processos pelos quais os dados passaram para chegar aos resultados foram registrados e devidamente documentados utilizando a biblioteca Python PROV-PY², utilizada para a criação de documentos de proveniência que implementa o modelo PROV-DM - W3C PROV Data Model, criado pela World Wide Web Consortium (W3C) [Belhajjame et al. 2013]. A documentação da proveniência de dados pode ser

²<https://pypi.org/project/prov/>

encontrada no repositório do projeto no GitHub³, que, por conter todas as etapas do fluxo de transformação dos dados, não poderá ser disposta em um formato gráfico neste artigo.

4.2. Coleta dos dados

Os dados do projeto foram providos pelo LASA, Laboratório de Aplicações de Satélites Ambientais do Departamento de Meteorologia da Universidade Federal do Rio de Janeiro (UFRJ), que tem suas atividades voltadas ao ensino e à pesquisa de técnicas de sensoria-mento remoto, aplicadas ao monitoramento da atmosfera e superfície terrestre.

A iniciativa ALARMES⁴ foi desenvolvida para servir como ferramenta de alerta rápido e ágil sobre o avanço da extensão da área afetada pelo fogo de forma a apoiar os órgãos ambientais nas ações de combate ao fogo. O sistema combina imagens de satélites da NASA, focos de calor e inteligência artificial para identificar novas áreas atingidas pelo fogo, monitorando diariamente a localização e extensão das áreas queimadas, permitindo, por exemplo, entender a velocidade de aumento dessas áreas afetadas. Essas informações são enviadas aos órgãos competentes para o planejamento e a realizações de ações efetivas. Neste contexto foi escolhido o bioma Pantanal para objeto de estudo, tendo em vista sua grande importância para o suprimento de água, a estabilização do clima e a conservação do solo no país.

Os datasets utilizados foram disponibilizados em diversos formatos (.shp, .tif, .dbf, .prj, .shx), particulares ao universo de dados geoespaciais, e contém informações sobre as queimadas no bioma Pantanal no Brasil. Primeiramente contemplaram os anos de 2012 a 2020 e depois foram obtidos dados na extensão .tif e dados com as mesmas extensões, do ano de 2021, tornando possível, assim, realizar uma análise exploratória da última década. O dataset, após concatenações, possui 346.666 linhas e 6 colunas.

Os dados obtidos possuem os seguintes atributos, conforme os metadados disponibilizados pelo LASA:

- Dia juliano: dia do ano em que a área foi afetada — Tipo de dado: int
- Mês: mês em que a área foi afetada — Tipo de dado: int
- Ano: ano em que a área foi afetada — Tipo de dado: int
- Geometry: coordenadas geográficas da área afetada em formato de polígono — Tipo de dado: geometry
- Confiança: confiança do sistema em relação à área afetada — Tipo de dado: int
- Versão: versão do sistema — Tipo de dado: str

4.3. Análise exploratória dos dados

Essa seção apresenta a análise exploratória dos dados (EDA), que consiste em uma etapa importante do processo de Ciência de Dados, pois permite que o pesquisador obtenha informações sobre os dados antes de realizar quaisquer suposições, tratamentos ou modelagens. Dessa forma é possível identificar erros, outliers, valores faltantes, distribuição dos dados, além de encontrar padrões e tendências que possam ser utilizados para prever os dados futuros.

³https://github.com/charlespimentel/DS_queimadas_pantanal/blob/main/prov_generator.ipynb

⁴<https://alarmes.lasa.ufrj.br/>

A EDA seguiu 3 etapas de análise dos dados e, para cada uma delas, foram utilizados conjuntos de bibliotecas Python específicas para análises geoespaciais, como o GeoPandas, Pandas, Matplotlib, Seaborn e Plotly. As etapas foram:

- Informações gerais do dataset: número de linhas, colunas, tipos de dados, valores nulos, etc.
- Transformação dos dados: transformação dos dados para o formato de dados geoespaciais, para ser possível realizar a análise espacial;
- Análises gráficas e inferências: produção de gráficos e análises estatísticas para identificar padrões e tendências nos dados.

Abaixo, é possível visualizar as principais ações realizadas e percepções obtidas a partir das análises gráficas e inferências.

Após o levantamento das estatísticas descritivas, houve uma necessidade de transformar os dados de geometria para compreensão de área de queimada. Para isso, foi utilizada a biblioteca GeoPandas, que permite a manipulação de dados geoespaciais, e a biblioteca Shapely, que permite a manipulação de dados de geometria. A transformação foi realizada utilizando a função `to_crs()`, que converte os dados de geometria para o sistema de coordenadas Universal Transverse Mercator (UTM), o qual é um sistema de coordenadas geográficas projetadas, com origem no meridiano de Greenwich e no equador, sendo utilizado para representar a superfície terrestre em mapas. Após a transformação, foi possível calcular a área de cada polígono, utilizando a função `area`, e adicionar essa informação como uma nova coluna no dataset.

A transformação dos dados de geometria para o sistema de coordenadas UTM possibilitou a realização de análises espaciais. A visualização dos dados em um mapa viabilizou a identificação de padrões e tendências, como a concentração de focos de calor em determinadas regiões, a variação da área afetada ao longo do ano, entre outros. A Figura 3 abaixo apresenta um mapa com a área afetada por focos de calor no bioma Pantanal nos últimos 10 anos.

Mapa de focos de queimadas no Pantanal de 2012 à 2021

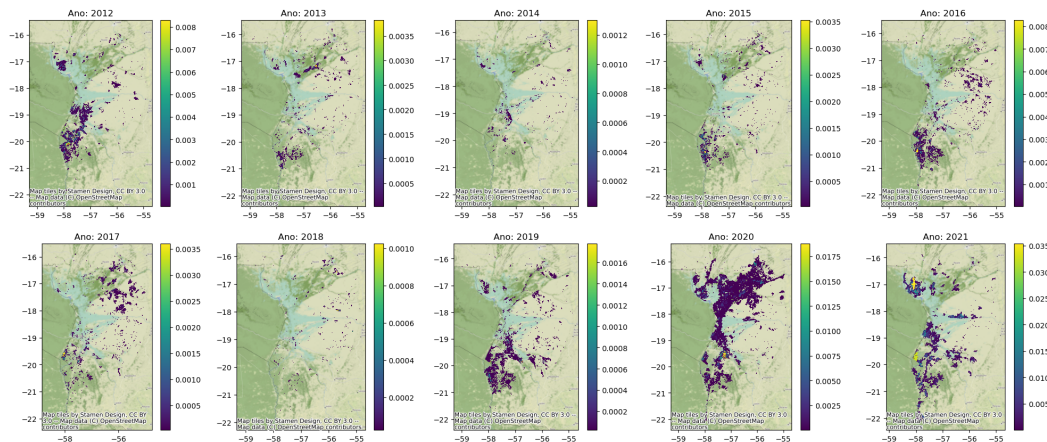


Figura 3. Mapa de focos de queimadas no Pantanal de 2012 à 2021.

Observa-se claramente que a área afetada por focos de queimadas aumentou ao longo dos anos, com um pico imenso em 2020, e que a maioria dos focos de calor

concentra-se na região central do bioma, em Mato Grosso do Sul. Os motivos para o crescimento da área afetada são diversos, mas podem ser explicados, principalmente, pela expansão da agricultura e pecuária na região, atividades que utilizam fogo para limpeza de pastagens e plantações, e pela falta de fiscalização e de políticas públicas para o controle do fogo. De 2019 para 2020, o bioma enfrentou uma seca histórica, que contribuiu para o cenário de incêndios florestais, e que foi agravada pela pandemia de COVID-19, que afetou a capacidade de resposta dos órgãos de fiscalização e de combate aos incêndios, como observado na Figura 4 abaixo.

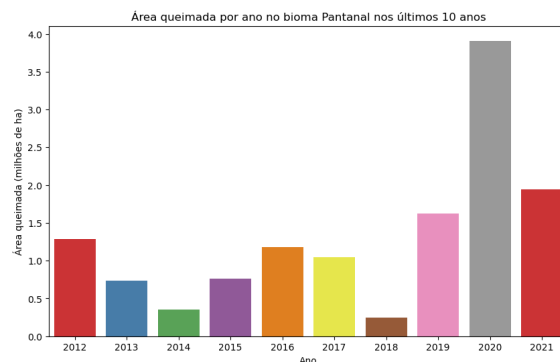


Figura 4. Mapa de focos de queimadas no Pantanal dos últimos 10 anos.

Para determinar a recorrência de queimadas em área mapeadas, foi realizada uma análise de sobreposição utilizando os dados de foco de calor e geometria do bioma Pantanal. O mapa foi construído utilizando a interseção, que permite a identificação de áreas que se sobrepõem entre conjuntos de dados geoespaciais a partir de atributos temporais.

Mapa de focos de queimadas no Pantanal de 2012 à 2021

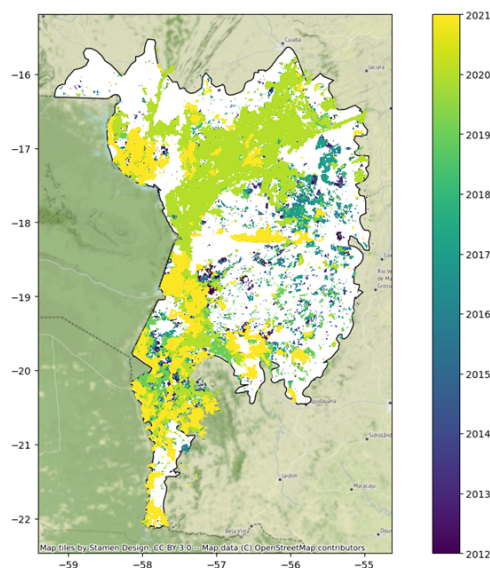


Figura 5. Mapa de focos de queimadas no Pantanal dos últimos 10 anos.

Com base nas informações levantadas, foi possível gerar uma tabela que compreende o total de área afetada por ano e suas respectivas porcentagens em relação ao total

de área do bioma Pantanal. A Tabela 1 abaixo apresenta os resultados.

Tabela 1. Porcentagem de área queimada em relação à área total do Pantanal no Brasil

Ano	Área queimada (Milhões de ha)	Percentual
2020	3,91	25,34%
2021	1,95	12,61%
2019	1,62	10,52%
2012	1,28	8,32%
2016	1,18	7,63%
2017	1,05	6,80%
2015	0,77	4,96%
2013	0,73	4,75%
2014	0,35	22,60%
2018	0,25	16,00%

No trabalho buscou-se identificar áreas de preservação e observar o impacto das queimadas no bioma Pantanal. Para isso, foi utilizado o dataset de áreas protegidas do Brasil, disponibilizado pelo IBGE. Para simplificação do dataset, foram utilizadas apenas as áreas protegidas pertencentes ao Sesc, entidade privada que possui a maior área de preservação no Brasil. Combinando a área protegida tem-se, aproximadamente, 88 mil hectares. A Figura 6 apresenta o mapa com a área protegida e a área afetada por focos de calor no bioma Pantanal com sobreposição temporal.

Mapa de focos de queimadas no Pantanal de 2012 à 2021 com as RPPNs Sesc Pantanal

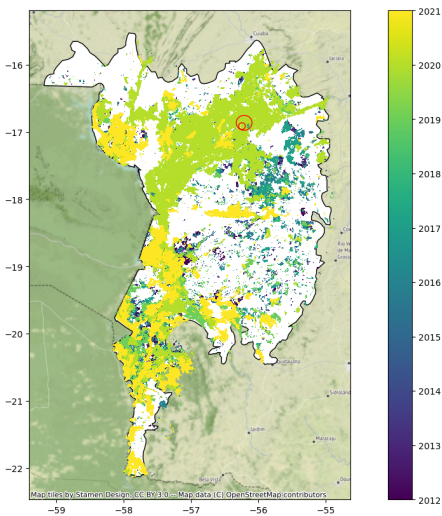


Figura 6. Mapa com a área protegida e a área afetada por focos de calor no Pantanal.

4.4. Preparação dos dados

Para essa seção, foi necessário realizar uma nova iteração do processo de ciência de dados, que desta vez, foi focado na modelagem da solução. Anteriormente o processo foi concentrado na coleta, tratamento e exploração dos dados brutos, disponibilizados pelos pesquisadores do LASA.

Primeiramente, foi necessário criar um dataset contendo somente datas e suas respectivas totalizações de áreas queimadas, tendo em vista o objetivo de modelagem preditiva. Logo após, foi realizada uma nova análise exploratória dos dados, que consistiu na realização de análises estatísticas e gráficas para identificação de possíveis correlações entre as variáveis, verificação da qualidade dos dados e mapeamento de tendências. A Figura 7 apresenta um gráfico de linha com a evolução da área afetada por focos de calor no bioma Pantanal ao longo dos anos.

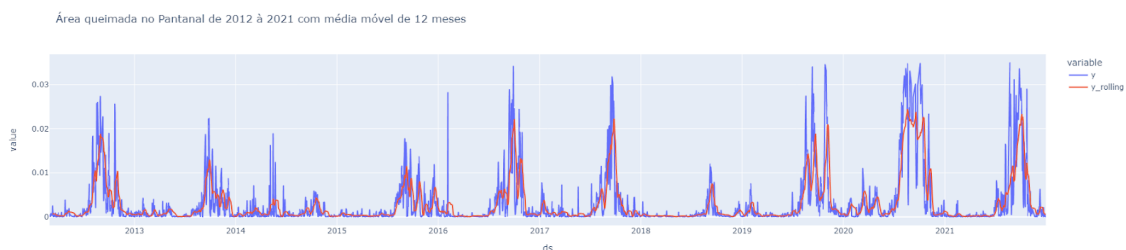


Figura 7. Gráfico de linha com a evolução da área afetada por focos de calor no Pantanal.

Através disso, percebeu-se uma quantidade considerável de outliers, ou seja, valores que estão muito distantes da média, conforme indicado na Figura 8. Para remover esses outliers, foi calculado o desvio padrão e a média dos dados fornecidos. Os valores que estavam fora do intervalo de 3 desvios padrão foram removidos.

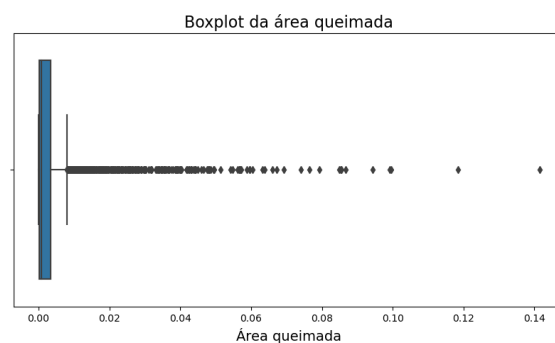


Figura 8. Diagrama de caixa da distribuição de área queimada.

Ao remover os outliers, podemos ter um modelo preditivo mais preciso, pois os dados que estão muito distantes da média podem ser considerados ruídos sem representar a realidade. Valores negativos também foram removidos, pois não fazem sentido para o problema e não existem no dataset original.

4.5. Análise preditiva

Pôde-se observar na análise exploratória realizada uma tendência de aumento no número de queimadas ao longo dos 10 últimos anos. Existem períodos específicos em que os focos de incêndio são mais frequentes, como no verão e no início do outono. Além disso, é possível observar que o número de queimadas aumenta em anos de seca, como 2019 e 2020. Neste contexto, um dos objetivos do projeto é viabilizar formas atenuar os efeitos das queimadas através de informações que possam auxiliar no processo de tomada de decisão.

Atualmente a área de aprendizagem de máquina é uma das mais promissoras para a solução de problemas complexos. Através de técnicas e algoritmos especializados é possível prever o comportamento de um sistema, como, por exemplo, o número de queimadas em um determinado período. Tendo em vista a importância do tema, o projeto buscou desenvolver um modelo de aprendizagem de máquina que seja capaz de prever o número de queimadas para o ano de 2022. Para isso, foi utilizado um conjunto de algoritmos e bibliotecas de aprendizagem de máquina, que analisam os dados históricos de queimadas para executar a predição e criar um modelo preditivo replicável.

Após a etapa de preparação do dataset histórico, foi realizada a modelagem preditiva (Figura 9). O modelo foi desenvolvido utilizando o Prophet, que é uma biblioteca python de código aberto para previsão de séries temporais. Criada pela área de pesquisas em Inteligência Artificial da Meta (Anteriormente Facebook), a biblioteca é baseada em um modelo preditivo aditivo, baseado em algoritmos de regressão linear. O modelo consegue capturar tendências não lineares, sazonalidade e feriados em dados de séries temporais, além de possuir uma API simples e intuitiva. Foram realizadas parametrizações para melhor ajustar o modelo aos dados e avaliações visando pontuar métricas de desempenho.

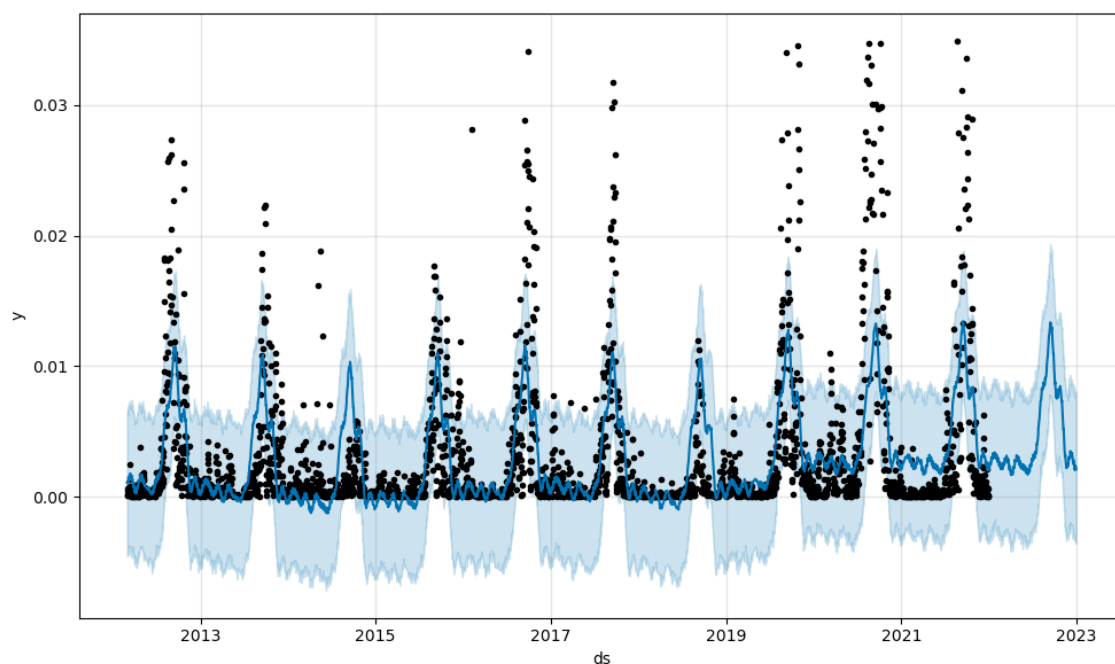


Figura 9. Gráfico de predição de valores para 2022 gerado pelo modelo preditivo.

4.6. Análise comparativa

Nesta etapa o intuito é comparar os resultados obtidos com o modelo preditivo com os dados reais de queimadas. Para isso, seria necessário obter os dados reais de queimadas para o ano de 2022. Porém, esses dados não foram disponibilizados pelos pesquisadores do LASA, tendo em vista que até a data de publicação deste artigo, os dados reais de queimadas para o ano de 2022 ainda não estavam disponíveis. Ademais, o LASA disponibiliza em sua Plataforma ALARMES um conjunto de informações consolidadas sobre o monitoramento de queimadas, que inclui dados de queimadas em tempo real, dados históricos de queimadas, dados de satélite, entre outros. Porém, esses dados são disponibilizados em visualizações interativas, que não permitem a extração dos dados em formato aberto.

5. Princípios FAIR

F.A.I.R. é um acrônimo que significa *Findable, Accessible, Interoperable and Reusable*. Publicado em 2016 por [Wilkinson et al. 2016], os princípios FAIR visam garantir que os dados sejam fáceis de serem encontrados, acessados, intercambiados e reutilizados. Para isso, é necessário que os dados sejam descritos de forma clara e completa, que possuam um identificador único, que sejam acessíveis de um ponto de vista operacional e que possuam uma licença que permita a reutilização.

Neste trabalho as diretrizes FAIR foram aplicadas para garantir a reprodutibilidade científica. Para isso, foi criado um repositório no GitHub⁵, uma plataforma de hospedagem de código com controle de versão usando o Git. O repositório contém o código-fonte do projeto, que pode ser acessado por qualquer pessoa que tenha interesse em utilizar o modelo preditivo. Além disso, o repositório contém um arquivo README, que descreve o projeto e fornece instruções para a instalação e utilização do modelo. O repositório também contém um arquivo LICENSE, que especifica a licença de uso do projeto, o qual foi licenciado sob a licença MIT, que permite a utilização do código-fonte para fins comerciais e não comerciais, além de permitir a modificação e redistribuição. O trabalho também foi publicado no Zenodo, uma plataforma de repositório de dados científicos, que permite a publicação de conjuntos de dados e códigos-fonte.

Em relação aos ambiente de desenvolvimento, foram utilizados o Google Colab, Anaconda e Docker. O Google Colab é um ambiente de desenvolvimento gratuito disponibilizado pelo Google que permite a execução de código em Python, R e SQL. O Anaconda é uma distribuição de código aberto e gratuita de Python e R, que permite a criação de ambientes virtuais complexos e compatíveis para a execução de projetos. O Docker é uma plataforma de código aberto que assegura a criação de máquinas e ambientes virtuais para produção e desenvolvimento de software. Sendo assim, o projeto foi desenvolvido em um ambiente de desenvolvimento colaborativo e reprodutível, que permite a execução do código em qualquer computador.

6. Resultados

Análise de dados ambientais promovem um conjunto amplo de possibilidades de geração de conhecimento, visto que são de grande importância para a sociedade, essenciais para a tomada de decisões e para a elaboração de políticas públicas. O projeto buscou viabilizar

⁵https://github.com/charlespimentel/DS_queimadas_pantanal

diferentes produtos de pesquisa que possam contribuir para a prevenção e mitigação de queimadas e incêndios no bioma Pantanal no Brasil.

6.1. Novo dataset de formato acessível

A área de pesquisa geoespacial lida frequentemente com dados de sensoriamento remoto, dados que possuem uma grande quantidade de informações latentes, mas que não são facilmente interpretáveis por humanos. Para isso, se faz necessário o uso de programas específicos que interpretam os dados e geram informações que podem ser utilizadas para a tomada de decisão. Um exemplo de programa que interpreta dados de sensoriamento remoto é o QGIS, Sistema de Informação Geográfica livre e aberto. O LASA processa, analisa e exporta os dados dos satélites ambientais aplicando técnicas de monitoramento da atmosfera e superfície terrestres.

Geralmente os dados exportados possuem formatos especificamente voltados para softwares de sensoriamento remoto, como o GeoTIFF (Figura 10), um formato de dados geoespaciais disposto em uma grade de pixels em uma imagem, dificultando uma análise estatística.



Figura 10. Mapa de queimadas de 2020 dispostas em imagem de formato GeoTiff.

Após análise e transformação dos dados no experimento, foi gerado um dataframe com as informações de queimadas no bioma Pantanal no Brasil, que possui um formato mais adequado para a análise exploratória dos dados e para a visualização dos resultados para um público não técnico da área de Ciências Geoespaciais. O formato de exportação do dataframe foi o .geojson, formato de dados geoespaciais aberto e de fácil visualização,

seja em softwares de SIG, como QGIS, ou em plataformas de visualização de dados, como o Google Earth Engine. Também pode ser consumido por diversas bibliotecas de visualização de dados, como o geopandas, uma biblioteca de código aberto para análise de dados geoespaciais em Python. Para disponibilizar ainda mais formatos abertos para a exploração, o dataset também foi exportado para o formato .csv, amplamente utilizado em ciências de dados.

Os novos datasets gerados podem ser utilizados por pesquisadores e estudantes de outras áreas do conhecimento para a análise de dados de queimadas no bioma Pantanal servem de base para a criação de novos experimentos e projetos de pesquisa em outros biomas brasileiros.

6.2. Predição de queimadas para o ano de 2022

Treinado com base nos dados históricos fornecidos anteriormente, o modelo mencionado na seção de metodologia foi utilizado para prever a área de queimada para o ano de 2022. O resultado final foi um dataset com a previsão de área queimada para cada dia do ano do ano, que pôde ser explorado através de uma visualização interativa no dashboard e que está disponível como fonte de dados aberta para download no repositório do projeto. O modelo criado também foi disponibilizado como arquivo .json, que pode ser facilmente importado por métodos do Prophet.

6.3. Dashboard dinâmico

A fim de facilitar a visualização dos resultados para um público não técnico da área de Ciências Geoespaciais e buscando ratificar a importância de proteger o bioma Pantanal, foi desenvolvido um dashboard dinâmico utilizando a biblioteca Streamlit, uma biblioteca de código aberto para a criação de aplicativos web para a linguagem de programação Python.

O dashboard foi desenvolvido utilizando o dataset de focos de queimadas do Pantanal de 2012 a 2021, que foi previamente tratado e analisado e possui 3 páginas, sendo elas: a página inicial, que apresenta um mapa com a localização dos focos de queimadas no Pantanal, a página de análise exploratória dos dados, que apresenta gráficos com informações sobre os dados de queimadas no Pantanal, e a página de previsão de focos de queimadas, que apresenta um gráfico com a previsão de focos de queimadas no Pantanal para o ano de 2022. O arquivo de código do dashboard está disponível no repositório do projeto no GitHub.

O dashboard traz uma visão mais ampla sobre a situação atual do Pantanal e contribui para a conscientização da população acerca da importância da proteção ambiental e da prevenção e combate a queimadas no bioma.

7. Considerações finais e trabalhos futuros

O projeto foi desenvolvido com o objetivo de promover a geração de conhecimento e a disseminação de informações sobre o impacto das queimadas no bioma Pantanal no Brasil. Apesar de ser o menor bioma brasileiro, o Pantanal é um dos mais importantes ecossistemas do país, considerado um dos maiores ecossistemas úmidos do mundo. As queimadas promovem uma devastação ambiental e uma perda irreversível de biodiversidade, além de causar impactos socioeconômicos e culturais.

É sabido que existem diversos fatores que contribuem para as queimadas, como a ação humana, a seca, o aquecimento global e a ação de incêndios controlados. O projeto teve como base dados históricos de áreas queimadas, mas acredita-se que a análise de outros fatores também possa contribuir para a compreensão do problema e para a criação de soluções para a prevenção e combate. Portanto, como trabalhos futuros, considera-se que uma análise mais holística dos dados de queimadas a partir de outros fatores e seus respectivos conjuntos de dados possa fundamentar a criação de soluções cada vez mais efetivas e completas.

Referências

- Belhajjame, K., B'Far, R., Cheney, J., Coppens, S., Cresswell, S., Gil, Y., Groth, P., Klyne, G., Lebo, T., McCusker, J., et al. (2013). Prov-dm: The prov data model. *W3C Recommendation*, 14:15–16.
- Correa, D. B., Alcântara, E., Libonati, R., Massi, K. G., and Park, E. (2022). Increased burned area in the pantanal over the past two decades. *Science of The Total Environment*, 835:155386.
- Leal Filho, W., Azeiteiro, U. M., Salvia, A. L., Fritzen, B., and Libonati, R. (2021). Fire in paradise: Why the pantanal is burning. *Environmental Science & Policy*, 123:31–34.
- Li, X., Song, K., and Liu, G. (2020). Wetland fire scar monitoring and its response to changes of the pantanal wetland. *Sensors*, 20(15):4268.
- Libonati, R., DaCamara, C. C., Peres, L. F., Sander de Carvalho, L. A., and Garcia, L. C. (2020). Rescue brazil's burning pantanal wetlands.
- Marques, J. F., Alves, M. B., Silveira, C. F., e Silva, A. A., Silva, T. A., Dos Santos, V. J., and Calijuri, M. L. (2021). Fires dynamics in the pantanal: Impacts of anthropogenic activities and climate change. *Journal of Environmental Management*, 299:113586.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., et al. (2016). The fair guiding principles for scientific data management and stewardship. *Scientific data*, 3(1):1–9.
- Wirth, R. and Hipp, J. (2000). Crisp-dm: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, volume 1, pages 29–39. Manchester.