

CAMPOS ALAGADOS E ÁREAS PANTANOSAS COMO BIOINDICADORES DE PERIGO DE INCÊNDIO NO CERRADO: UMA ABORDAGEM DE MACHINE LEARNING

Iuri Gabriel Meris¹
Heloisa Cristine Hahn²
Celso Voos Viera³

RESUMO

Este estudo investigou o potencial dos campos alagados e áreas pantanosas como bioindicadores de perigo de incêndio no Cerrado, utilizando o Índice de Vegetação por Diferença Normalizada (NDVI) e a Temperatura da Superfície Terrestre (LST) como variáveis-chave. Foram coletados dados de 20 pontos amostrais localizados em formação pantanosa (Coleção 9 do MapBiomias), todos com recorrência mínima de 20 queimas registradas de 1985 a dezembro de 2023, conforme a Coleção 3 do MapBiomias Fogo. Para avaliar a correlação entre o NDVI, a LST e as áreas queimadas mensais, aplicaram-se técnicas de aprendizado de máquina (Regressão Linear Simples, Regressão Linear Múltipla e K-Nearest Neighbors Regressor – KNN), com uso das métricas R^2 e MSE para mensurar o desempenho preditivo. Os resultados mostraram que a Regressão Linear Múltipla explicou 74,37% da variância nas áreas queimadas no conjunto de teste, enquanto o KNN ($k = 9$) atingiu 81,35%, configurando-se como o método com melhor desempenho preditivo. Conclui-se que a vegetação de campos alagados e áreas pantanosas atua como bioindicador das condições propícias a incêndios, servindo como ferramenta metodológica no monitoramento ambiental e no planejamento de políticas públicas de prevenção no Cerrado.

PALAVRAS-CHAVE: Perigo de incêndio; NDVI; LST; aprendizado de máquina; áreas úmidas.

FLOODED GRASSLANDS AND MARSH AREAS AS BIOINDICATORS OF FIRE DANGER IN THE CERRADO: A MACHINE LEARNING APPROACH

ABSTRACT

This study investigated the potential of flooded grasslands and marsh areas as bioindicators of fire danger in the Cerrado, using the Normalized Difference Vegetation Index (NDVI) and Land Surface Temperature (LST) as key variables. Data were collected from 20 sampling points located in marsh formations, according to MapBiomias Collection 9, all of which presented a minimum recurrence of 20 fire events recorded between 1985 and December 2023, according to MapBiomias Fire Collection 3. To assess the relationship between NDVI, LST, and monthly burned area, machine learning techniques were applied, including Simple Linear Regression, Multiple Linear Regression, and the K-Nearest Neighbors Regressor (KNN). The R^2 and mean squared error (MSE) metrics were used to evaluate predictive performance. The results showed that Multiple Linear Regression explained 74.37% of the variance in burned area in the test dataset, while the KNN model with $k = 9$ achieved 81.35%, representing the best predictive performance among the evaluated methods. The findings indicate that flooded grasslands and marsh vegetation can act as bioindicators of conditions conducive to fire occurrence, providing a methodological tool for environmental monitoring and the planning of public policies for fire prevention in the Cerrado.

KEYWORDS: Fire danger; NDVI; LST; machine learning; wetlands.

¹ Biólogo. MBA em Inteligência de Dados Ambiental. Joinville, Santa Catarina, Brasil; iuri.meris@gmail.com

² Engenheira química. Joinville, Santa Catarina, Brasil; heloisacristinehc@gmail.com

³ Doutor em Geografia. Professor e pesquisador do Programa de Pós-Graduação em Saúde e Meio Ambiente da Universidade da Região de Joinville – Univille, Joinville, Santa Catarina, Brasil; celso.v@univille.br

CAMPOS INUNDABLES Y ÁREAS PANTANOSAS COMO BIOINDICADORES DEL PELIGRO DE INCENDIO EN EL CERRADO: UN ENFOQUE DE APRENDIZAJE AUTOMÁTICO

RESUMEN

Este estudio investigó el potencial de los campos inundables y las áreas pantanosas como bioindicadores del peligro de incendio en el Cerrado, utilizando el Índice de Vegetación de Diferencia Normalizada (NDVI) y la Temperatura de la Superficie Terrestre (LST) como variables clave. Se recopilieron datos de 20 puntos de muestreo ubicados en formaciones pantanosas, según la Colección 9 de MapBiomias, todos ellos con una recurrencia mínima de 20 eventos de quema registrados entre 1985 y diciembre de 2023, de acuerdo con la Colección 3 de MapBiomias Fuego. Para evaluar la relación entre el NDVI, la LST y el área quemada mensual, se aplicaron técnicas de aprendizaje automático, incluidas la Regresión Lineal Simple, la Regresión Lineal Múltiple y el Regresor de K Vecinos Más Cercanos (KNN). Se utilizaron las métricas R^2 y error cuadrático medio (MSE) para evaluar el desempeño predictivo. Los resultados mostraron que la Regresión Lineal Múltiple explicó el 74,37 % de la varianza del área quemada en el conjunto de prueba, mientras que el modelo KNN con $k = 9$ alcanzó el 81,35 %, presentando el mejor desempeño predictivo entre los métodos evaluados. Los resultados indican que la vegetación de los campos inundables y de las áreas pantanosas puede actuar como bioindicadora de condiciones propicias para la ocurrencia de incendios, constituyendo una herramienta metodológica para el monitoreo ambiental y la planificación de políticas públicas de prevención de incendios en el Cerrado.

PALABRAS CLAVE: Peligro de incendio; NDVI; LST; aprendizaje automático; áreas húmedas.

1.INTRODUÇÃO

O Cerrado, o segundo maior bioma da América do Sul, cobre uma área de aproximadamente 2.036.448 km², representando cerca de 22% do território brasileiro. Sua abrangência inclui áreas contínuas em estados como Goiás, Tocantins, Mato Grosso, Mato Grosso do Sul, Minas Gerais, Bahia, Maranhão, Piauí, Rondônia, Paraná, São Paulo e o Distrito Federal, além de encaves em Amapá, Roraima e Amazonas (MMA, 2012). Reconhecido mundialmente por sua biodiversidade, o Cerrado é frequentemente descrito como a “caixa d’água do Brasil”, devido à sua importância na manutenção de recursos hídricos. No entanto, esse bioma enfrenta pressões crescentes, especialmente durante a estação seca, quando a alteração dos regimes naturais de fogo pode intensificar impactos ecológicos, socioambientais e econômicos.

Segundo o MapBiomias Fogo, a Coleção 4 apresenta o mapeamento anual e mensal de cicatrizes de fogo no Brasil no período de 1985 a 2024, incluindo área queimada anual, área queimada mensal, frequência, área acumulada, tamanho das cicatrizes e ano da última ocorrência de fogo (MAPBIOMAS, 2025). Em 2024, o Brasil registrou cerca de 30 milhões de hectares queimados, valor 62% superior à média histórica. Nesse contexto, o Cerrado destacou-

se como o segundo bioma com maior extensão afetada pelo fogo, com 10,6 milhões de hectares queimados, correspondendo a 35% do total nacional e a um crescimento de 10% em relação à média histórica do próprio bioma, estimada em 9,6 milhões de hectares por ano. Entre 1985 e 2024, o Cerrado acumulou 89,5 milhões de hectares queimados ao menos uma vez, equivalente a aproximadamente 45% de sua área, evidenciando a relevância desse bioma para estudos sobre perigo de incêndio, recorrência de fogo e Manejo Integrado do Fogo (MAPBIOMAS, 2025).

Embora o fogo tenha desempenhado papel fundamental na evolução da flora do Cerrado, promovendo adaptações morfológicas e ecológicas específicas (SIMON et al., 2009), sua interpretação deve considerar que o Cerrado é um ecossistema dependente do fogo. Nesses ambientes, determinados regimes de queima contribuíram historicamente para a manutenção de espécies, habitats e paisagens; entretanto, alterações na frequência, intensidade e sazonalidade do fogo podem comprometer a resiliência ecológica e ampliar os efeitos negativos sobre a biodiversidade e os serviços ecossistêmicos (MYERS, 2006; PIVELLO, 2011). Assim, alterações antropogênicas, como a expansão agrícola, o desmatamento e o manejo inadequado do solo, têm modificado os regimes naturais de fogo, ampliando a extensão das áreas queimadas e os impactos ambientais associados (ARRUDA et al., 2021).

As áreas de Campo Alagado e Pantanosa, caracterizadas por vegetação predominantemente herbácea adaptada ao alagamento sazonal ou permanente, desempenham um papel singular no bioma Cerrado. Segundo o MapBiomas (Coleção 9, 2024), essas formações cobrem cerca de 42% das áreas alagadas do Brasil, totalizando 61 milhões de hectares, e apresentam as maiores frequências de queima no bioma, com recorrência de até 26 queimas em algumas regiões. Essas áreas são especialmente suscetíveis ao fogo devido à presença de biomassa seca durante a estação seca, que atua como material combustível. Além disso, mudanças climáticas globais e a conversão antrópica do uso do solo têm intensificado esses eventos, ampliando os impactos ambientais e as emissões de gases de efeito estufa associados às áreas queimadas (ARRUDA et al., 2021).

A compreensão dos padrões espaço-temporais de ocorrência de fogo no Cerrado é essencial para o desenvolvimento de estratégias eficazes de prevenção e manejo. Tecnologias de sensoriamento remoto, como o uso do Índice de Vegetação por Diferença Normalizada (NDVI) e da Temperatura da Superfície Terrestre (LST), têm se mostrado ferramentas valiosas para monitorar o vigor vegetativo, as condições biofísicas da cobertura vegetal e os impactos das áreas queimadas (LIMA et al., 2013; CHEN et al., 2016). Esses índices permitem identificar

áreas prioritárias para conservação e avaliar a resposta da vegetação a eventos de fogo, contribuindo para uma gestão mais eficiente dos recursos naturais.

Recentemente, avanços em aprendizado de máquina têm ampliado as possibilidades de análise de dados ambientais, permitindo o desenvolvimento de modelos preditivos capazes de estimar padrões associados à ocorrência de áreas queimadas (MOHAMMED et al., 2020). A integração de algoritmos de aprendizado de máquina com dados geoespaciais e climáticos têm potencial para melhorar o planejamento territorial e a implementação de políticas públicas de manejo do fogo. Além disso, essa abordagem contribui para a construção de políticas territoriais mais éticas e eficientes, ao integrar variáveis socioeconômicas e ambientais.

Neste contexto, o presente estudo propôs investigar como as áreas de Campo Alagado e Pantanosa do Cerrado podem atuar como bioindicadores de perigo de incêndio. Utilizando técnicas de aprendizado de máquina aplicadas ao NDVI e à LST, buscou-se compreender a relação entre variáveis biofísicas e áreas queimadas mensais nessas formações. O objetivo é fornecer subsídios científicos para ações de monitoramento e prevenção, contribuindo para a conservação deste importante bioma e para o manejo sustentável de seus recursos.

2. METODOLOGIA

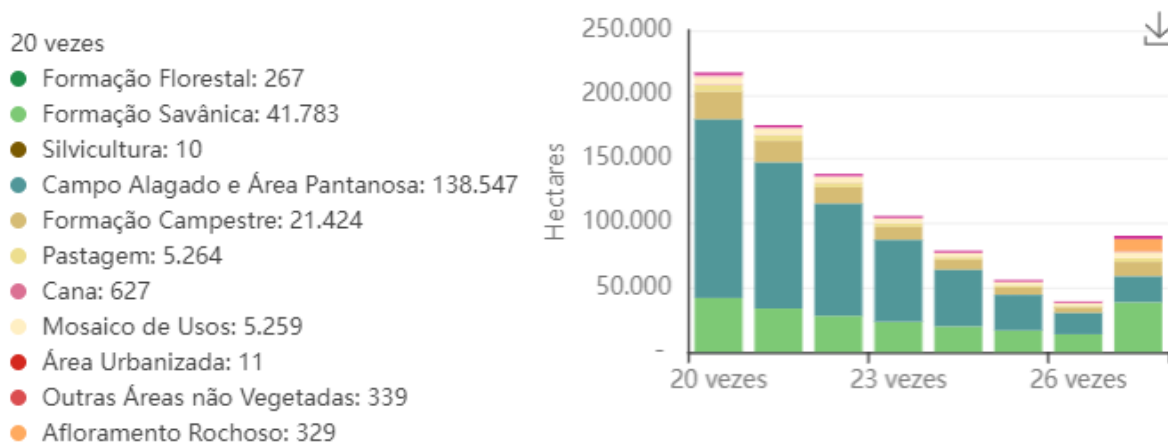
2.1 ÁREA DE ESTUDO

O estudo foi conduzido no bioma Cerrado, que abrange cerca de 22% do território brasileiro e apresenta grande diversidade de ecossistemas, incluindo formações de Campos Alagados e Áreas Pantanosas. Este tipo de vegetação foi escolhido devido à sua alta suscetibilidade à ocorrência de fogo e à relevância como potencial bioindicador de incêndios florestais.

Para o estudo, foram selecionados 20 pontos de amostragem localizados em áreas classificadas como Formação Pantanosa, conforme os dados do Coleção 9 do MapBiomas (MapBiomas, 2024). Os pontos foram escolhidos com base na alta frequência de queimadas, com registros superiores a 20 eventos de 1985 a 2022. A maioria dos pontos concentrou-se no estado de Mato Grosso, complementados por amostras no Tocantins e Maranhão.

A Figura 1 ilustra a frequência de áreas queimadas em diferentes classes de cobertura e uso do solo no Cerrado, enquanto a Figura 2 apresenta a localização geográfica dos pontos de amostragem.

Figura 1 - Frequência de área queimada por classe de cobertura e uso do solo.



Fonte: MapbiomasFogo, 2024.

Figura 2 - Mapa de localização dos pontos de amostragem.

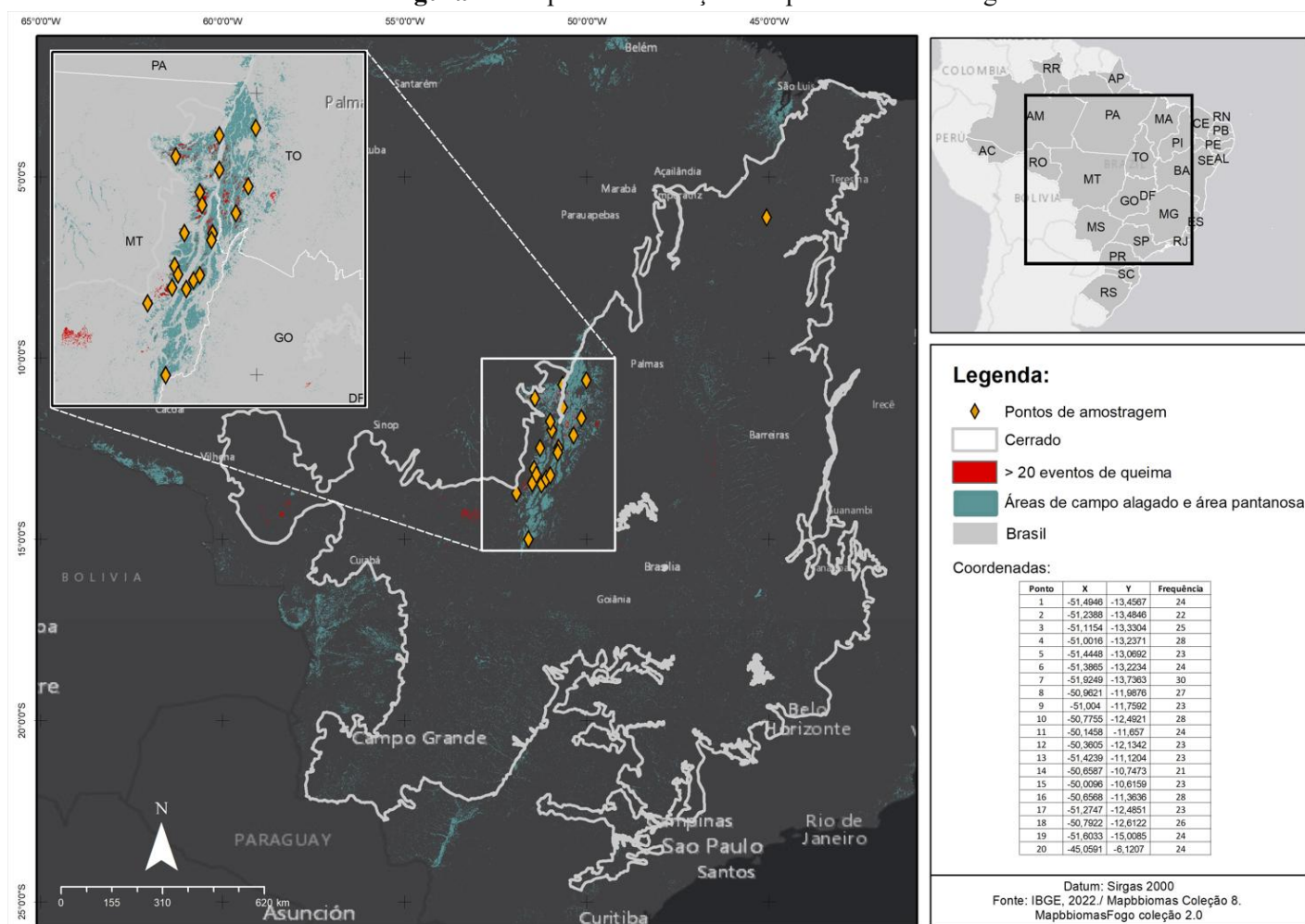


Tabela 1 — Coordenadas geográficas, estado, classe MapBiomias e recorrência de fogo (1985–2022) dos 20 pontos amostrais.

Ponto	Latitude	Longitude	Estado	Classe MapBiomias	Recorrência (1985–2022)
1	-13,4567°	-51,4946°	MT	Formação Pantanosa	24
2	-13,4846°	-51,2388°	MT	Formação Pantanosa	22
3	-13,3304°	-51,1154°	MT	Formação Pantanosa	25
4	-13,2371°	-51,0016°	MT	Formação Pantanosa	28
5	-13,0692°	-51,4448°	MT	Formação Pantanosa	23
6	-13,2234°	-51,3865°	MT	Formação Pantanosa	24
7	-13,7363°	-51,9249°	MT	Formação Pantanosa	30
8	-11,9876°	-50,9621°	TO	Formação Pantanosa	27
9	-11,7592°	-51,0040°	TO	Formação Pantanosa	23
10	-12,4921°	-50,7755°	TO	Formação Pantanosa	28
11	-11,6570°	-50,1458°	TO	Formação Pantanosa	24
12	-12,1342°	-50,3605°	TO	Formação Pantanosa	23
13	-11,1204°	-51,4239°	TO	Formação Pantanosa	23
14	-10,7473°	-50,6587°	TO	Formação Pantanosa	21
15	-10,6159°	-50,0096°	TO	Formação Pantanosa	23
16	-11,3636°	-50,6568°	TO	Formação Pantanosa	28
17	-12,4851°	-51,2747°	TO	Formação Pantanosa	23
18	-12,6122°	-50,7922°	TO	Formação Pantanosa	26
19	-15,0085°	-51,6033°	MT	Formação Pantanosa	24
20	-6,1207°	-45,0591°	MA	Formação Pantanosa	24

Fonte: elaboração própria. Coordenadas em sistema geodésico SIRGAS 2000. Recorrência conforme camada de frequência da Coleção 3.0 do MapBiomias Fogo. Classe MapBiomias conforme Coleção 9 (MapBiomias, 2024).

2.2 COLETA DE DADOS

Todos os dados utilizados nesta pesquisa derivam de produtos de sensoriamento remoto operados por instituições oficiais, garantindo padronização metodológica e reprodutibilidade. Os dados foram coletados em escala mensal, abrangendo o período de fevereiro de 2000 a dezembro de 2023, para cada um dos 20 pontos amostrais descritos na Tabela 1.

Com relação ao Índice de Vegetação por Diferença Normalizada (NDVI), os valores foram obtidos por meio da plataforma SATVeg (SATVeg, 2016) da Embrapa Agricultura

Digital, que disponibiliza séries temporais derivadas do produto MOD13Q1 do sensor MODIS (NASA), com resolução espacial de 250 metros e composição temporal de 16 dias. O NDVI é um indicador do vigor fotossintético da vegetação, sensível à biomassa verde e à umidade foliar, sendo amplamente utilizado em estudos de monitoramento de ecossistemas e detecção de áreas suscetíveis a incêndios.

Quanto à Temperatura da Superfície Terrestre (LST), os dados foram extraídos do produto MOD11A2 do sensor MODIS (NASA), com resolução espacial de 1 quilômetro e composição de 8 dias (WAN; HOOK; HULLEY, 2021), por meio da plataforma Google Earth Engine. A LST representa a temperatura radiométrica da superfície e, em ambientes savânicos, é um indicador da disponibilidade hídrica do sistema solo-vegetação-atmosfera, relacionando-se com o secamento do material combustível e com a probabilidade de ignição.

Os dados de área queimada mensal foram obtidos da Coleção 3.0 do MapBiomas Fogo, que mapeia cicatrizes de queimadas no território brasileiro a partir de imagens Landsat com 30 metros de resolução espacial, processadas por algoritmos de aprendizado profundo. A escolha dessa fonte, em detrimento de produtos de detecção térmica como focos de calor, justifica-se pela natureza da variável modelada: o estudo busca compreender a ocorrência e propagação do fogo a partir de cicatrizes pós-evento, e não a detecção em tempo real. A variável foi expressa em hectares de área queimada por mês, agregada para o conjunto dos 20 pontos amostrais.

Essa abordagem analítica busca compreender a ocorrência e propagação do fogo nesses ambientes, subsidiando um sistema de alerta para antecipar e reduzir incêndios florestais no bioma.

2.3 PRÉ-PROCESSAMENTO DE DADOS

2.3.1 CORREÇÕES E FILTRAGEM

Para assegurar a integridade dos dados de NDVI, foi realizado o processo de pré-filtragem para eliminar inconsistências oriundas de cobertura de nuvens e outros artefatos visuais. Este procedimento envolveu o uso de interpolação bilinear para remover pontos de dados inválidos (NoData) e corrigir valores distorcidos pela presença das nuvens, utilizando os metadados de qualidade de pixel disponibilizados pela própria plataforma SATVeg.

Os dados de LST, originalmente em Kelvin (K), foram convertidos para graus Celsius (°C).

2.3.2 AGREGAÇÃO TEMPORAL

Os valores coletados foram consolidados em médias mensais para cada ponto de amostragem. Isso permitiu comparar diretamente os índices vegetativos, as temperaturas e os registros de área queimada ao longo do período estudado.

2.3.3 NORMALIZAÇÃO

As variáveis foram normalizadas para o intervalo [0, 1] aplicando-se o método Min-Max Scaling, implementado pela classe MinMaxScaler do módulo sklearn.preprocessing (PEDREGOSA et al., 2011), conforme a expressão:

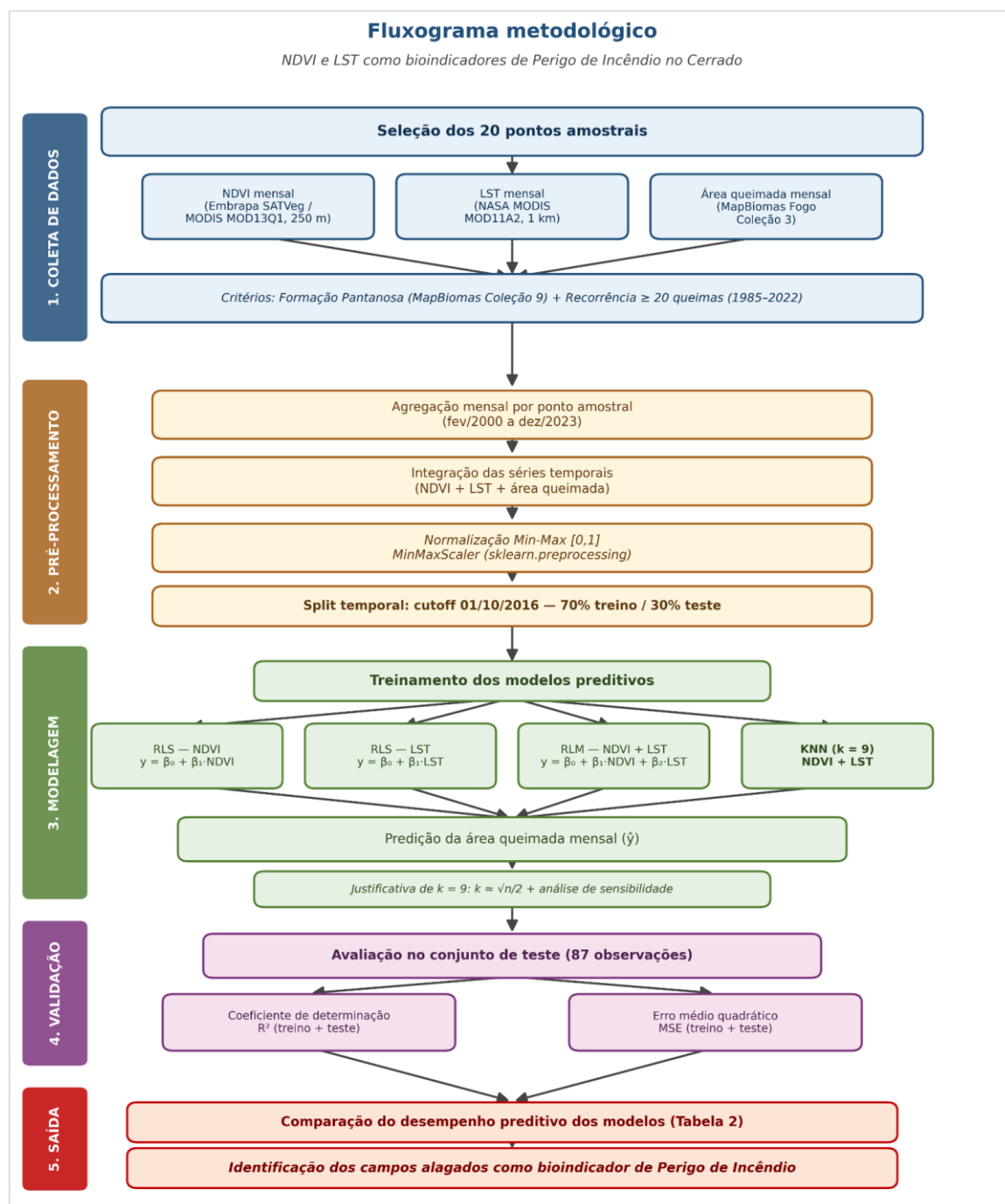
$$x_{norm} = (x - x_{min}) / (x_{max} - x_{min})$$

Esse procedimento foi essencial para equilibrar variáveis com escalas distintas, garantindo a compatibilidade com os algoritmos de aprendizado de máquina.

2.4 MODELAGEM E ANÁLISE ESTATÍSTICA

A Figura 3 apresenta, de forma sintética, o fluxo metodológico adotado, enquanto as subseções a seguir detalham cada etapa.

Figura 3 - Fluxograma metodológico



2.4.1 LINGUAGEM E FERRAMENTAS

A modelagem foi conduzida em linguagem Python, com uso das bibliotecas pandas (manipulação e organização das séries temporais), NumPy (operações numéricas em arrays multidimensionais), scikit-learn (PEDREGOSA et al., 2011) para a implementação dos algoritmos de aprendizado de máquina, e Matplotlib para a geração dos gráficos analíticos. A

escolha dessas ferramentas reflete a sua ampla adoção em estudos geoespaciais e ambientais, com suporte ativo, documentação técnica robusta e compatibilidade com pipelines de sensoriamento remoto.

2.4.2 ALGORITMOS E JUSTIFICATIVA METODOLÓGICA

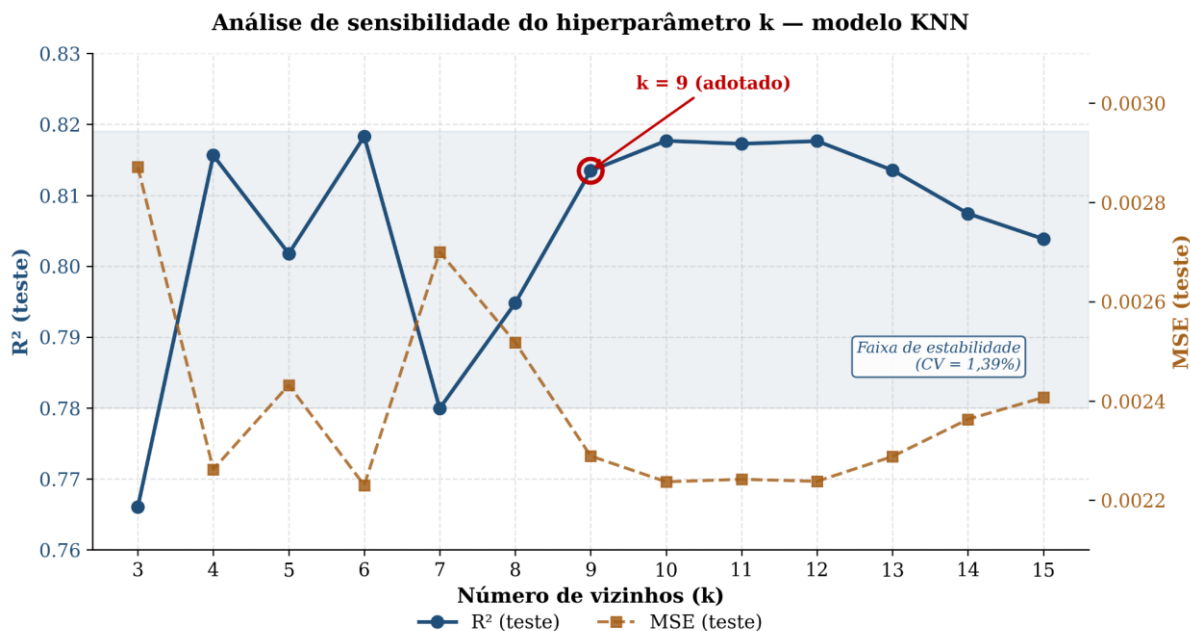
Para investigar a relação entre as variáveis biofísicas (NDVI e LST) e a área queimada no Cerrado, foram aplicados três tipos de modelos preditivos: Regressão Linear Simples (RLS), Regressão Linear Múltipla (RLM) e K-Nearest Neighbors Regressor (KNN). A Regressão Linear Simples foi aplicada de forma independente para cada variável explanatória, permitindo avaliar a contribuição isolada de cada índice na predição da área queimada. A Regressão Linear Múltipla combinou ambas as variáveis em um único modelo, ampliando a capacidade explanatória. O modelo KNN, de natureza não paramétrica, foi adotado para capturar relações não lineares entre as variáveis preditoras e a variável-resposta, atribuindo a cada amostra um valor de saída calculado como média dos k vizinhos mais próximos no espaço de atributos normalizado.

O conjunto de dados, composto por 287 observações mensais (fevereiro de 2000 a dezembro de 2023), foi dividido em conjunto de treinamento e conjunto de teste mediante split temporal, com data de corte em 01 de outubro de 2016 — resultando em 200 observações para treino (70%) e 87 para teste (30%). A escolha pelo split temporal, em detrimento da divisão aleatória, justifica-se pela natureza das séries temporais: o embaralhamento de observações causaria vazamento temporal, permitindo que o modelo observasse o futuro durante o treinamento. A separação cronológica preserva a integridade preditiva da avaliação.

O hiperparâmetro $k = 9$ do modelo KNN foi definido a partir de análise de sensibilidade, realizou-se análise de sensibilidade testando valores de k de 3 a 15 (Figura 4), na qual o desempenho do modelo manteve-se estável no intervalo $k \in [4, 15]$, com coeficiente de variação do R^2 de 1,39%. Esse resultado confirma a robustez da escolha e indica que o desempenho do KNN não é sensível a pequenas variações no hiperparâmetro, no intervalo investigado.

O desempenho dos modelos foi avaliado pelas métricas coeficiente de determinação (R^2) e erro médio quadrático (MSE), aplicadas tanto ao conjunto de treino quanto ao conjunto de teste. Os resultados quantitativos são apresentados na Seção 3.

Figura 4 - Sensibilidade ao k.

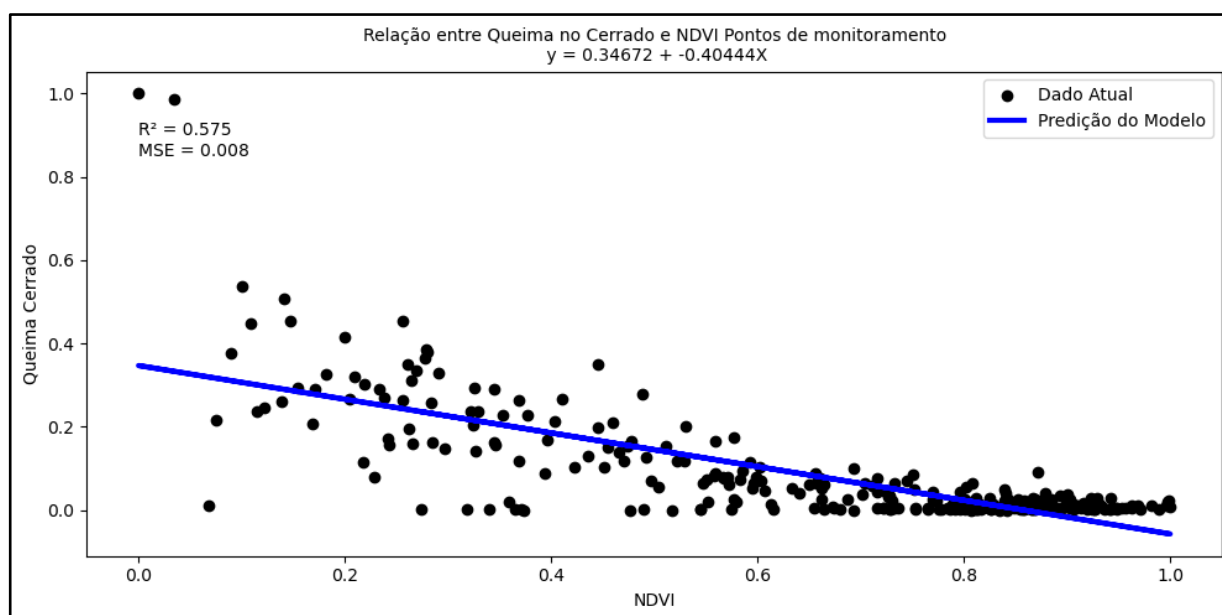


3. RESULTADOS

3.1 REGRESSÃO LINEAR (RL) SIMPLES

Na análise de regressão simples, ambas as variáveis, NDVI e LST, demonstraram uma moderada capacidade de explicação com índices de determinação (R^2) substanciais, sendo 0,575 para o NDVI (Figura 5) e 0,667 para o LST (Figura 6). Adicionalmente, ambas as variáveis apresentaram um erro quadrático médio (MSE) abaixo de 0,01, indicando uma boa precisão dos modelos na previsão das áreas queimadas no cerrado.

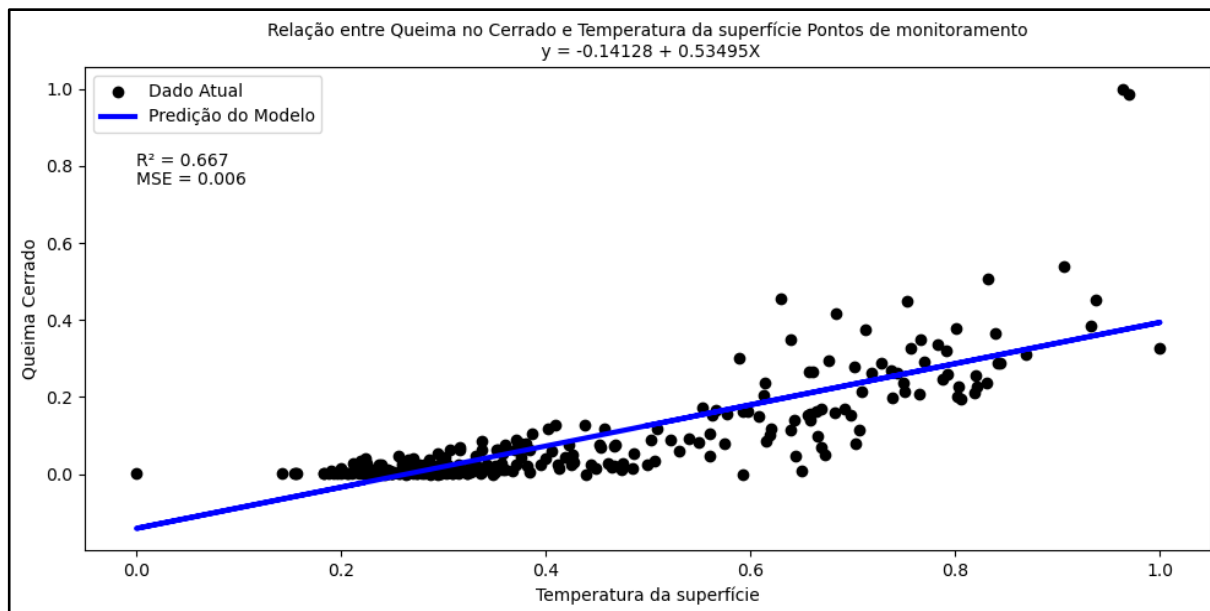
Figura 5 - Gráfico de Regressão linear simples Área queimada no Cerrado x NDVI.



Equação 1: Relação entre Área queimada no Cerrado x NDVI Pontos de amostragem

$$y = 0,34672 + -0,40444X$$

Figura 6 - Gráfico de Regressão linear simples Área queimada no Cerrado x LST.



Equação 2: Relação entre Área queimada no Cerrado x LST Pontos de amostragem

$$y = -0,14128 + 0,53495X$$

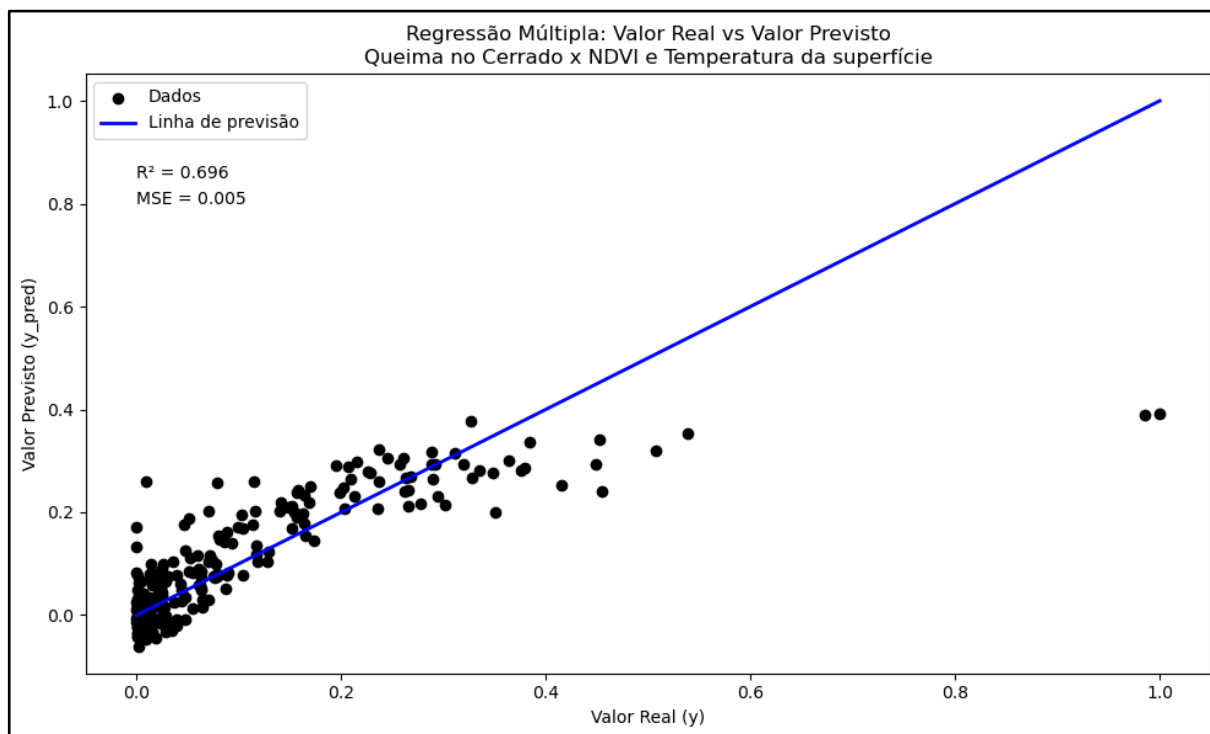
3.2 REGRESSÃO LINEAR (RL) MÚLTIPLA

Na análise de regressão múltipla, observou-se que a combinação das variáveis independentes NDVI (Índice de Vegetação por Diferença Normalizada) e LST (Temperatura da Superfície Terrestre) resultou em um aumento substancial no coeficiente de determinação (R^2), que alcançou o valor de 0,696.

Este resultado indicou que, em conjunto, essas variáveis conseguiram explicar 69,6% da variância na área queimada total no bioma do Cerrado.

Além disso, o modelo demonstrou alta precisão, conforme evidenciado pelo baixo Erro Médio Quadrático (MSE) de 0,005, indicando que as previsões do modelo estão muito próximas dos valores reais observados.

Figura 7 - Gráfico de regressão linear múltipla Área queimada no Cerrado x NDVI x LST.



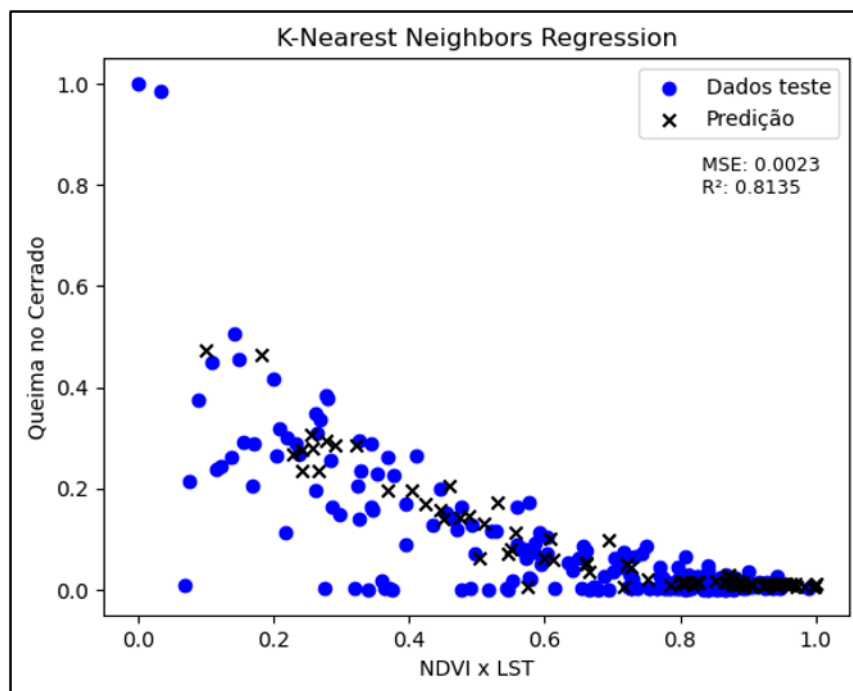
Equação 3: Relação entre Área queimada no Cerrado x NDVI Pontos de amostragem x LST Pontos de amostragem.

$$y = 0,021433 + 0,38403 * l_{st} + -0,15249 * ndvi$$

3.3 K NEIGHBORS REGRESSOR (KNN)

Utilizando o algoritmo KNN com as mesmas variáveis independentes (NDVI e LST), houve um incremento no coeficiente de determinação (R^2), que atingiu o valor de 0,8135, conforme pode ser observado na Figura 8. Isso sugere que a combinação dessas variáveis pode explicar 81,35% da variância na área queimada total no bioma do Cerrado.

A precisão do modelo é reforçada ainda mais através do Erro Médio Quadrático (MSE) apresentando um valor baixo, registrado em 0,002, o que reafirma que as estimativas do modelo estão consistentemente próximas dos valores reais.

Figura 8 - Gráfico K-NearestRegressor Área queimada no Cerrado x NDVI x LST.

3.4 TESTE E VALIDAÇÃO

A divisão da amostra em dados de treinamento e de teste é realizada para verificar se um modelo apresenta boa performance não apenas em dados utilizados para seu ajuste (treinamento), mas também capacidade de generalização para novas observações (teste). (SANTOS et al., 2019).

O período considerado para os treinamentos e testes dos modelos abrange um histórico de 287 meses (02/2000 a 12/2023) de queimadas.

Para validação dos modelos de predição (RL e KNN) realizou-se a divisão dos dados em amostra e treinamento de acordo com a ordem cronológica. As proporções dos dados foram 70% (200 meses) para treinamento do modelo e 30% (87 meses) para teste e avaliação.

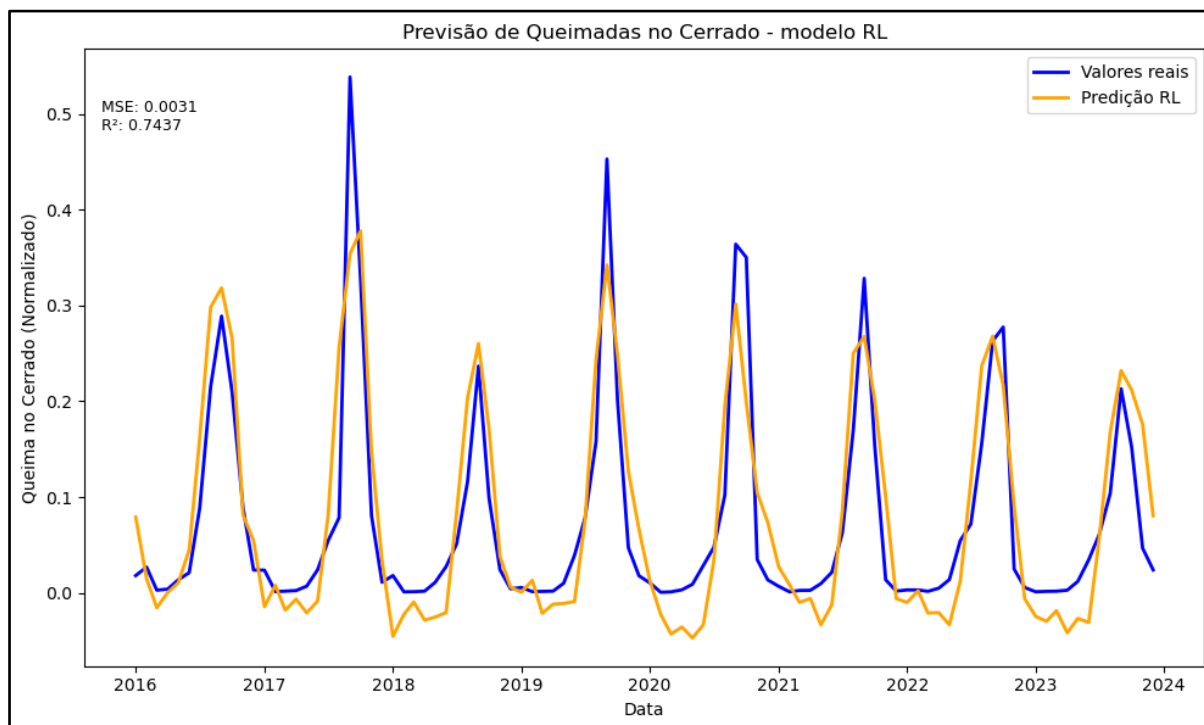
Após o treinamento, os modelos foram submetidos a um teste prático utilizando o conjunto de dados separado para essa finalidade (X_{test}).

As previsões geradas (y_{pred}) foram então comparadas com os valores reais (y_{test}), o que permitiu avaliar a qualidade e precisão das previsões dos modelos.

3.5 REGRESSÃO LINEAR

A análise comparativa entre a regressão linear simples e a regressão linear múltipla indicou um desempenho superior do modelo múltiplo, que, por sua vez, foi exclusivamente utilizado nas fases de teste e validação.

Figura 9 - Teste e validação do modelo.



De acordo com a Figura 9, é possível observar que a avaliação do modelo no teste produziu um Erro Médio Quadrático (MSE) de 0,0031 e um coeficiente de determinação (R^2) de 0,7437, evidenciando um ajuste satisfatório do modelo, apesar da ocorrência de valores negativos em determinados períodos.

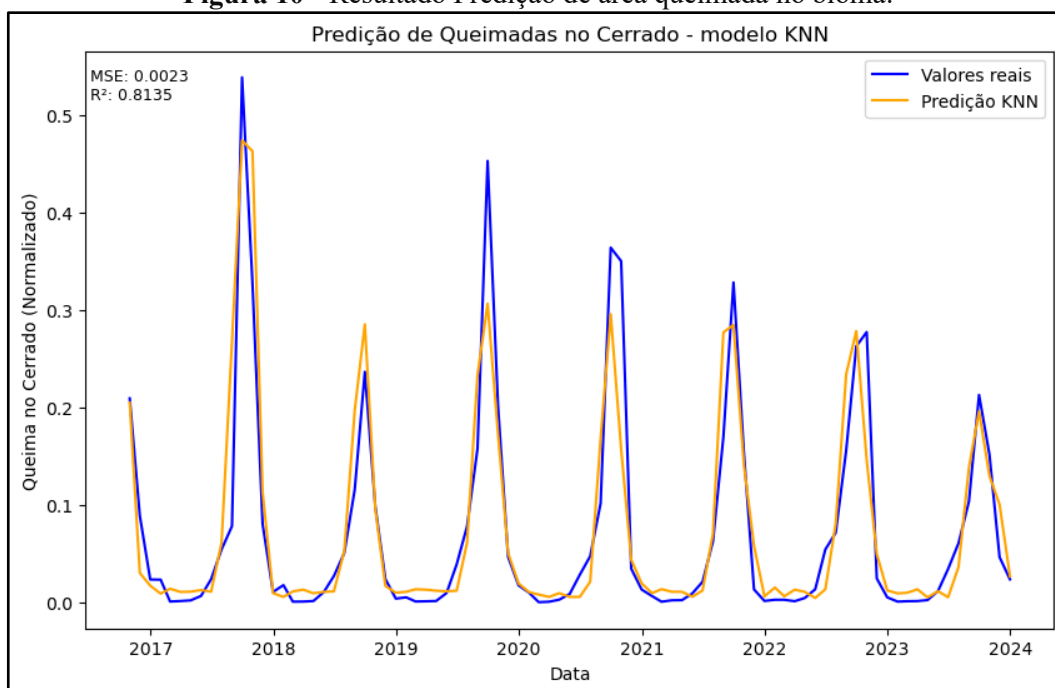
O baixo MSE sugeriu que os erros de previsão são, em média, reduzidos, enquanto o R^2 de 74,37% demonstrou que o modelo conseguiu explicar uma considerável parcela da variação na área queimada no Cerrado, com base nas variáveis NDVI e LST dos 20 pontos amostrais das áreas de campos alagados e áreas pantanosas.

3.6 K-NEAREST REGRESSOR (KNN)

Os resultados obtidos com o modelo KNN mostraram uma melhoria em relação à regressão linear múltipla, conforme evidenciado na Figura 10.

No teste, o modelo KNN alcançou um MSE de 0,0023 e um R^2 de 0,8135, evidenciando que este modelo apresentou capacidade superior de prever com precisão a área queimada no Cerrado.

Este desempenho revelou a robustez do KNN em capturar as dinâmicas complexas associadas às variáveis ambientais analisadas.

Figura 10 - Resultado Predição de área queimada no bioma.

A Tabela 2 sintetiza o desempenho dos modelos avaliados, contemplando tanto a análise descritiva sobre o conjunto total de observações (Seções 3.1 a 3.3) quanto a validação cronológica do RLM e do KNN sobre o conjunto de teste (Seções 3.5 e 3.6). Observa-se incremento progressivo do coeficiente de determinação à medida que se avança da regressão simples para a múltipla, e da regressão paramétrica para a regressão por vizinhança (KNN), confirmando que a interação não linear entre NDVI e LST é mais bem capturada pelo modelo KNN.

Tabela 2 — Síntese comparativa dos modelos avaliados. Modelos descritivos (RLS NDVI, RLS LST, RLM) ajustados sobre o conjunto total ($n = 287$). Validação cronológica (cutoff 01/10/2016, 70% treino / 30% teste) aplicada ao RLM validado e ao KNN.

Modelo	n	R ²	MSE	Abordagem
RLS — NDVI	287	0,575	0,008	Descritiva
RLS — LST	287	0,667	0,006	Descritiva
RLM — NDVI + LST	287	0,696	0,005	Descritiva
<i>Validação cronológica (cutoff 01/10/2016)</i>				
RLM — validada (teste)	87	0,7437	0,0031	Validação
KNN (k = 9) — validado (teste)	87	0,8135	0,0023	Validação

Fonte: elaboração própria. Variáveis normalizadas no intervalo [0, 1] (MinMaxScaler). Métricas adicionais do modelo KNN validado: MAE = 0,02709; RMSE = 0,04784. R²: coeficiente de determinação. MSE: erro médio quadrático.

4. DISCUSSÃO

Os dados analisados sugeriram uma relação inversamente proporcional entre o NDVI e as áreas queimadas no Cerrado, indicando que vegetação fotossinteticamente ativa estava associada a uma menor área queimada. Essa relação é representada pela equação 1, cuja pendente negativa refletiu a dinâmica observada. Estudos como os de Pereira Júnior (2002) e Prudente e Rosa (2016) já apontaram o NDVI como um indicador-chave para identificar áreas suscetíveis ao fogo, corroborando os resultados apresentados.

A relação inversa observada entre o NDVI e a área queimada não é apenas estatística, mas reflete um mecanismo fisiológico-ecológico bem documentado: o NDVI é sensível à biomassa fotossinteticamente ativa e ao conteúdo de água foliar, variáveis diretamente associadas à inflamabilidade do material combustível (YEBRA et al., 2013). Vegetação com elevado vigor fotossintético apresenta maior teor de umidade nos tecidos foliares, o que eleva a energia necessária para ignição e desfavorece a propagação do fogo. Inversamente, valores reduzidos de NDVI indicam senescência foliar ou perda de biomassa verde, condições que aumentam a proporção de combustível seco disponível e a probabilidade de ocorrência de queimadas. Em ambientes savânicos, essa relação é particularmente intensa durante a estação seca, quando a queda do NDVI antecede e prediz com razoável antecedência os pulsos de área queimada (CACCAMO et al., 2012; YEBRA et al., 2013).

Por outro lado, a Temperatura da Superfície Terrestre (LST) mostrou-se positivamente correlacionada com as áreas queimadas, conforme representado na equação 2. O coeficiente de determinação mais elevado ($R^2 = 0,667$) indicou que a LST é um preditor ligeiramente mais eficaz do que o NDVI isoladamente para esses eventos, sugerindo que o aumento das temperaturas favorece a propagação do fogo.

A correlação positiva entre LST e área queimada também encontra fundamento físico no processo de combustão. A temperatura da superfície terrestre é um proxy do balanço energético na interface solo-vegetação-atmosfera e atua como indicador indireto da disponibilidade hídrica do sistema: valores elevados de LST indicam reduzida evapotranspiração e perda de água no perfil superficial, condições que aceleram o secamento do material combustível e diminuem o ponto de ignição efetivo (HARDY, 2005). Em ambientes savânicos, esse mecanismo é amplificado durante a estação seca, quando a baixa precipitação e a alta radiação solar elevam a LST simultaneamente à redução do NDVI — caracterizando o que a literatura de ecologia do fogo denomina "janela de ignição" (HARDY, 2005). A

complementaridade entre NDVI (proxy da umidade vegetal) e LST (proxy da carga térmica) justifica o ganho preditivo observado quando ambas as variáveis são integradas em modelos multivariados, como ocorre no presente estudo.

Os campos alagados e formações pantanosas do Cerrado apresentam dinâmica peculiar quanto à acumulação de material combustível, o que ajuda a explicar sua elevada recorrência de fogo (média de 24 queimas no período 1985–2022 nos pontos amostrais analisados). Durante a estação chuvosa, o alagamento sazonal promove acumulação intensa de biomassa herbácea, sustentada pela alta disponibilidade hídrica e por adaptações fisiológicas das gramíneas dominantes. Com o recuo das águas no início da estação seca, essa biomassa entra em processo acelerado de senescência e secamento, resultando em uma camada espessa e contínua de material combustível fino — condições ideais para a propagação horizontal do fogo. Esse padrão fenológico contrasta com fitofisionomias savânicas adjacentes, onde a estratificação vertical do dossel e a presença de árvores com cascas espessas modulam o comportamento do fogo de forma distinta (PIVELLO, 2011; ALENCAR et al., 2022).

É importante ressaltar que o Cerrado é classificado ecologicamente como bioma dependente do fogo (fire-dependent), no qual o regime de queimadas naturais — disparadas predominantemente por raios no início da estação chuvosa — moldou ao longo de milhões de anos as adaptações morfológicas e fisiológicas da flora local (SIMON et al., 2009; PIVELLO, 2011). A persistência de espécies herbáceas e lenhosas com casca espessa, gemas subterrâneas e rebrota vigorosa após queimas evidencia coevolução com o fogo. Portanto, a interpretação dos resultados deste estudo deve considerar que área queimada não é, por si só, indicador de degradação: a alteração ecológica ocorre quando o regime natural de fogo é modificado em frequência, intensidade ou sazonalidade, especialmente pela substituição das queimas precoces e de baixa intensidade por incêndios tardios e severos durante o pico da estação seca (PIVELLO, 2011; SCHMIDT et al., 2018). Nesse contexto, o monitoramento do perigo de incêndio mediado por sensoriamento remoto cumpre função estratégica não apenas na prevenção de incêndios catastróficos, mas também na implementação do Manejo Integrado do Fogo (MIF) — política institucionalizada no Brasil desde 2014 (SCHMIDT et al., 2018) e regulamentada nacionalmente pela Lei 14.944/2024 (BRASIL, 2024).

Quando integrados em um modelo de regressão linear múltipla, NDVI e LST resultaram em um desempenho preditivo aprimorado ($R^2 = 0,696$), evidenciando que a interação entre as variáveis contribui para a compreensão e previsão das queimadas. Esse padrão foi reforçado no

modelo KNN, que apresentou um R^2 de 0,8135 e um MSE de 0,0023, destacando-se como a abordagem com melhor desempenho preditivo no presente estudo.

Os resultados deste estudo são compatíveis com a literatura recente sobre aplicação de aprendizado de máquina ao monitoramento de fogo em ambientes savânicos. Mohammed et al. (2020), em estudo comparativo de algoritmos para detecção precoce de incêndios florestais com dados geoespaciais, demonstraram a superioridade de métodos não paramétricos sobre regressões lineares para captura de relações complexas entre variáveis biofísicas e ocorrência de fogo. De forma análoga, Alencar et al. (2022) consolidaram o uso de redes neurais profundas para mapeamento de cicatrizes de queimadas a partir de séries temporais Landsat — base metodológica do MapBiomas Fogo Coleção 3 utilizado neste estudo. O desempenho preditivo do KNN observado nesta pesquisa ($R^2 = 0,8135$ no conjunto de teste) situa-se em faixa coerente com a literatura comparada, considerando que estudos com bases de dados mais amplas e variáveis adicionais (precipitação, velocidade do vento, déficit de pressão de vapor) tendem a alcançar coeficientes superiores. Esta convergência reforça a robustez metodológica da combinação NDVI + LST + KNN como ferramenta de baixo custo computacional para subsidiar sistemas operacionais de alerta de perigo de incêndio.

Esses achados reforçam o potencial de combinar variáveis ambientais para mapeamento de Perigo de Incêndios, com aplicações práticas na gestão de áreas protegidas e no planejamento de políticas públicas. Além disso, considerando que a maioria das áreas queimadas é causada por ações humanas, os resultados reforçam a necessidade de estratégias de mitigação, como educação ambiental, maior rigor legislativo e restrição de acesso a áreas vulneráveis.

A proximidade espacial observada entre alguns dos pontos amostrais reflete a distribuição natural das formações pantanosas no Cerrado, que ocorrem em manchas espacialmente contíguas, particularmente ao longo das planícies aluviais do rio Araguaia e seus afluentes. Trata-se, portanto, de uma amostragem dirigida pelo fenômeno de interesse, e não de viés metodológico: os critérios de seleção (Formação Pantanosa com recorrência de fogo superior a 20 queimas em 1985–2022) restringem o universo amostral a essas regiões específicas. O ponto 20, localizado na Chapada das Mesas (MA), foi propositalmente posicionado fora do núcleo amostral principal para avaliar a generalização espacial dos padrões observados em condições ecológicas distintas. A consistência dos resultados modelados, mesmo com essa amostra fora do núcleo, reforça a robustez da relação entre NDVI, LST e área queimada nas formações pantanosas do Cerrado.

Apesar dos resultados obtidos, este estudo apresenta limitações que merecem ser explicitadas. Primeiro, o modelo restringiu-se a duas variáveis biofísicas (NDVI e LST), não incorporando preditores climáticos diretamente associados ao regime de fogo, como precipitação acumulada, déficit de pressão de vapor (VPD) e velocidade do vento — variáveis amplamente reconhecidas como determinantes do comportamento do fogo (HARDY, 2005; YEBRA et al., 2013). Segundo, a análise não incluiu variáveis socioeconômicas ou de uso do solo no entorno dos pontos amostrais, fatores relevantes em um bioma cujos incêndios são majoritariamente de origem antrópica (PIVELLO, 2011). Terceiro, embora a validação cronológica preserve a integridade preditiva, ela não testa explicitamente a generalização espacial dos modelos para outras regiões do Cerrado fora das áreas de campos alagados e pantanosos. Estudos futuros poderão explorar a incorporação de variáveis climáticas e socioambientais em modelos integrados, ampliar a base amostral para outras fitofisionomias e avaliar arquiteturas de aprendizado profundo, como redes recorrentes (LSTM) ou modelos baseados em atenção, para captura de dependências temporais de longo prazo nas séries fenológicas.

Para além da contribuição metodológica, os resultados deste estudo dialogam diretamente com a agenda contemporânea de Manejo Integrado do Fogo (MIF) no Cerrado. A identificação operacional de áreas com elevado perigo de incêndio a partir de variáveis biofísicas acessíveis (NDVI e LST) pode subsidiar o planejamento de queimas prescritas precoces, ferramenta central do MIF para redução da carga de material combustível em fitofisionomias dependentes do fogo (SCHMIDT et al., 2018). A institucionalização do MIF no Brasil pela Lei 14.944/2024 (BRASIL, 2024) cria demanda crescente por sistemas de monitoramento operacionais, de baixo custo e com base científica robusta — nicho no qual abordagens como a apresentada neste estudo apresentam potencial direto de aplicação em unidades de conservação federais e estaduais do bioma.

5. CONCLUSÃO

Este estudo demonstrou o potencial dos campos alagados e áreas pantanosas como bioindicadores de perigo de incêndio no cerrado, com base em variáveis derivadas de sensoriamento remoto (NDVI e LST) e técnicas de aprendizado de máquina. A integração dessas variáveis em modelos preditivos mostrou-se cientificamente fundamentada: o NDVI reflete a umidade foliar e a disponibilidade de combustível verde, enquanto a LST atua como

indicador da carga térmica e do secamento do material combustível, configurando-se como preditores complementares da ocorrência de fogo.

Dentre os métodos testados, o modelo K-Nearest Neighbors (KNN, $k = 9$) apresentou o melhor desempenho preditivo, com R^2 de 81,35% e MSE de 0,0023 sobre o conjunto de teste, superando a regressão linear múltipla validada (R^2 de 74,37%). Esse resultado confirma a capacidade de modelos não paramétricos para captura de relações não lineares entre variáveis biofísicas e área queimada em ambientes savânicos, em consonância com a literatura recente sobre aplicação de aprendizado de máquina ao monitoramento de fogo.

Os resultados também ressaltam a especificidade ecológica das formações pantanosas do Cerrado: a dinâmica sazonal de alagamento e secamento favorece a acumulação intensa de biomassa herbácea, que se converte em material combustível durante a estação seca, explicando a elevada recorrência de fogo observada (média de 24 queimas no período 1985–2022). Reconhecer essas áreas como bioindicadores estratégicos amplia a precisão do monitoramento ambiental e fornece base científica para ações de prevenção, especialmente em um bioma classificado como dependente do fogo, no qual a alteração do regime natural — e não a presença do fogo em si — constitui a principal ameaça à conservação.

Do ponto de vista aplicado, a abordagem proposta tem potencial direto de integração ao Manejo Integrado do Fogo (MIF), política institucionalizada no Brasil desde 2014 e regulamentada nacionalmente pela Lei nº 14.944/2024. A identificação operacional de áreas com elevado perigo de incêndio a partir de variáveis biofísicas acessíveis pode subsidiar o planejamento de queimas prescritas precoces e o monitoramento contínuo em unidades de conservação federais e estaduais do bioma.

Como limitações, registra-se a restrição do modelo a duas variáveis biofísicas, sem incorporação de preditores climáticos diretos (precipitação, déficit de pressão de vapor, velocidade do vento) e socioeconômicos. Estudos futuros poderão ampliar a base preditiva com essas variáveis, testar arquiteturas de aprendizado profundo para captura de dependências temporais de longo prazo e avaliar a generalização espacial do modelo para outras fitofisionomias do Cerrado e biomas correlatos.

6. DISPONIBILIDADE DE DADOS E CÓDIGO

O código, os notebooks e os dados processados utilizados no pré-processamento, modelagem, validação e geração dos resultados deste estudo estão disponíveis no repositório:

<https://github.com/iurimeris/cerrado-wetlands-fire-ml>

REFERÊNCIAS

- ALENCAR, A. A. C. et al. Long-Term Landsat-Based Monthly Burned Area Dataset for the Brazilian Biomes Using Deep Learning. **Remote Sensing**, v. 14, n. 11, p. 2510, 2022. DOI: 10.3390/rs14112510.
- ARRUDA, V. L. S. et al. An alternative approach for mapping burn scars using Landsat imagery, Google Earth Engine, and Deep Learning in the Brazilian Savanna. **Remote Sensing Applications: Society and Environment**, v. 22, 2021.
- BRASIL. Lei nº 14.944, de 31 de julho de 2024. Institui a Política Nacional de Manejo Integrado do Fogo. Diário Oficial da União, Brasília, 2024.
- CACCAMO, G. et al. Assessing the sensitivity of MODIS to monitor drought in high biomass ecosystems. **Remote Sensing of Environment**, v. 116, p. 2754-2766, 2012.
- CHEN, W. et al. Mapping a burned forest area from Landsat TM data by multiple methods. **Geomatics, Natural Hazards and Risk**, v. 7, p. 384-402, 2016.
- HARDY, C. C. Wildland fire hazard and risk: problems, definitions, and context. **Forest Ecology and Management**, v. 211, n. 1-2, p. 73-82, 2005. DOI: 10.1016/j.foreco.2005.01.029.
- LIMA, G. C. et al. Avaliação da cobertura vegetal pelo índice de vegetação por diferença normalizada (IVDN). **Revista Ambiente & Água - An Interdisciplinary Journal of Applied Science**, v. 8, n. 2, 2013.
- MAPBIOMAS. Área queimada no Brasil em 2024 supera média histórica em 62%. MapBiomas Brasil, 24 jun. 2025. Acesso em: 13 maio 2026.
- MAPBIOMAS. Cerrado teve 639 mil hectares queimados em 2023. MapBiomas Fogo, 2023. Disponível em: <https://brasil.mapbiomas.org/>. Acesso em: 12 maio 2026.
- MAPBIOMAS. Mapeamento anual de cobertura e uso do solo no Brasil – Coleção 9. 2024. Disponível em: <https://brasil.mapbiomas.org/>. Acesso em: 12 maio 2026.
- MOHAMMED, Z. et al. Comparative study on machine learning algorithms for early fire forest detection system using geodata. **International Journal of Electrical and Computer Engineering**, v. 10, p. 5507-5513, 2020.
- MMA. Ministério do Meio Ambiente: O Bioma Cerrado, 2012. Disponível em: <https://antigo.mma.gov.br/biomas/cerrado.html>. Acesso em: 12 abr. 2024.
- MYERS, R. L. Convivendo com o fogo: manutenção dos ecossistemas e subsistência com o Manejo Integrado do Fogo. Tallahassee: The Nature Conservancy, 2006.
- PEDREGOSA, F. et al. Scikit-learn: Machine Learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825-2830, 2011.
- PEREIRA JÚNIOR, A. C. Métodos de geoprocessamento na avaliação da susceptibilidade do Cerrado ao fogo. 2002. 98 f. Tese (Doutorado em Ciências da área de Concentração em Ecologia e Recursos Naturais) - Universidade Federal de São Carlos, São Carlos, 2002.

PIVELLO, V. R. The Use of Fire in the Cerrado and Amazonian Rainforests of Brazil: Past and Present. **Fire Ecology**, v. 7, n. 1, p. 24-39, 2011. DOI: 10.4996/fireecology.0701024.

PRUDENTE, T.; ROSA, R. Análise dos incêndios florestais e variação temporal do NDVI no município de São Valério-TO. **Revista Formação (Online)**, v. 2, n. 23, p. 271-286, 2016.

SANTOS, H. J. et al. Machine learning para análises preditivas em saúde: exemplo de aplicação para prever óbito em idosos de São Paulo, Brasil. **Cadernos de Saúde Pública**, v. 35, 2019. Disponível em: <https://cadernos.ensp.fiocruz.br/static/arquivo/1678-4464-csp-35-07-e00050818.pdf>. Acesso em: 14 abr. 2024.

SATVeg. Sistema de Análise Temporal da Vegetação. Empresa Brasileira de Pesquisa Agropecuária, 2016. Disponível em: <https://www.satveg.cnptia.embrapa.br/satveg/login.html>. Acesso em: 14 abr. 2024.

SCHMIDT, I. B. et al. Fire management in the Brazilian savanna: First steps and the way forward. **Journal of Applied Ecology**, v. 55, n. 5, p. 2094-2101, 2018. DOI: 10.1111/1365-2664.13118.

SIMON, M. F. et al. Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by in situ evolution of adaptations to fire. **Proceedings of the National Academy of Sciences**, v. 106, n. 48, p. 20359-20364, 2009. Disponível em: <https://doi.org/10.1073/pnas.0903410106>. Acesso em: 5 nov. 2024.

WAN, Z.; HOOK, S.; HULLEY, G. MODIS/Terra Temperatura/Emissividade da superfície terrestre 8 dias L3 Global 1km SIN Grid V061, 2021. Disponível em: <https://doi.org/10.5067/MODIS/MOD11A2.061>. Acesso em: 14 abr. 2024.

YEBRA, M. et al. A global review of remote sensing of live fuel moisture content for fire danger assessment: Moving towards operational products. **Remote Sensing of Environment**, v. 136, p. 455-468, 2013. DOI: 10.1016/j.rse.2013.05.029.