

Método para a Classificação de Áreas Queimadas Baseado em Aprendizado de Máquina Automatizado

Marcelly Homem Coelho

Programa de Pós-Graduação em Computação Aplicada
Instituto Nacional de Pesquisas Espaciais – INPE
São José dos Campos, São Paulo - Brasil
marcellyhc@gmail.com

Olga Oliveira Bittencourt

Laboratório de Computação Aplicada
Instituto Nacional de Pesquisas Espaciais – INPE
São José dos Campos, São Paulo - Brasil
olgarf.oliveira@gmail.com

Fabiano Morelli

Programa Queimadas
Instituto Nacional de Pesquisas Espaciais – INPE
São José dos Campos, São Paulo - Brasil
fabiano.morelli@inpe.br

Rafael Santos

Programa de Pós-Graduação em Computação Aplicada
Instituto Nacional de Pesquisas Espaciais – INPE
São José dos Campos, São Paulo - Brasil
rafael.santos@inpe.br

ABSTRACT

Forest fires burn large areas of native vegetation and it causes impacts in the social, economic and ecological scope. Burnt areas classification can help understand fires occurrence and support public policies. This work aims to develop a method of automatic burnt areas classification. The method is based on the application of Automated Machine Learning in data sets, from the Landsat-8/OLI satellite images of 2018 and 2019. We intend to answer the following research question: “*Is it possible to automate the choice of machine learning models and maintain quality levels in the classification of burnt areas?*”. The contribution of this research is to determine whether a predictive model, trained with validated samples from 2018, is capable of classifying fires occurrences in 2019. For the performance evaluation, the following metrics were analyzed: precision, probability of detection and average success rate. The results indicate that the method has a high potential to classify burnt areas.

KEYWORDS

Burnt areas, Automated machine learning, Classification, Supervised learning, Remote sensing

1 INTRODUÇÃO

De acordo com dados estatísticos divulgados pelo Programa Queimadas do Instituto Nacional de Pesquisas Espaciais (INPE), nas últimas décadas, tem aumentado as áreas de queimadas no Brasil. Os seis biomas brasileiros sofreram danos nos seus ecossistemas florestais e tiveram parte da sua cobertura vegetal perdida. Entre janeiro a dezembro de 2020, da relação total de focos ativos, a porcentagem de incêndios ativos detectados nos biomas foram: Amazônia (46,3%), Cerrado (28,6%), Pantanal (9,9%), Mata Atlântica (7,9%), Caatinga (6,6%), e Pampa (0,7%). Dos países que integram a América do Sul, no mesmo período, o Brasil

ocupou a primeira posição em quantidade de queimadas, representando 48,5% do total de focos ativos detectados [2].

Estudos apontam que os efeitos das queimadas nos ecossistemas são complexos, abrangendo desde a perda de biodiversidade da fauna e flora e a redução da fertilidade do solo a impactos ecológicos devido sua influência na poluição atmosférica e nas mudanças climáticas. Além disso, representam uma fonte adicional de emissões de gases de efeito estufa e podem resultar em diversas implicações, como paralisação de atividades aeroportuárias, aumento de morbimortalidade por doenças respiratórias, danos ao patrimônio público e privado, dentre outras [6][17].

Os avanços tecnológicos na área espacial possibilitaram o desenvolvimento de sensores remotos a bordo de satélites, os quais são capazes de capturar imagens da Terra que abrangem áreas amplas e apresentam alta frequência temporal [3][5]. Esta realidade favoreceu o aprimoramento de técnicas de sensoriamento remoto, visto que, as imagens constituem importante fonte de informação que podem ser usadas para o monitoramento de incêndios florestais.

A detecção e o monitoramento de incêndios florestais são fundamentais para viabilizar o controle do fogo e a redução dos custos nas operações de combate às queimadas. De acordo com [8], usar dados de satélite para promover a classificação automática de áreas queimadas de forma rápida, contínua e confiável é um grande desafio, devido a necessidade de um conjunto de dados de conhecimento. Segundo [9], muitas regiões do Brasil têm poucos conjuntos de dados validados por especialistas.

Em [8][9], os autores destacam a diversidade de modelos de aprendizado de máquina que podem ser aplicados para classificar áreas queimadas. Este fator torna a escolha manual de um algoritmo uma tarefa difícil. Novas tecnologias estão surgindo para resolver este problema. Conforme [11], as ferramentas de geração de Aprendizado de Máquina Automatizado (do Inglês, *Automated Machine Learning*) (AutoML) podem selecionar modelos e suas respectivas configurações de hiperparâmetros que apresentem o

melhor desempenho nos conjuntos de dados de entradas fornecidos ao algoritmo de treinamento.

Neste contexto, este artigo tem como objetivo avaliar um método para a classificação automática de áreas queimadas, do bioma Cerrado, baseado na aplicação de AutoML. Para esse fim, foram usadas amostras de queimadas, validadas por especialistas, referentes aos anos de 2018 e 2019. O objetivo principal da pesquisa consiste em determinar se um modelo preditivo, treinado com amostras de 2018, é capaz de classificar ocorrências de queimadas de 2019. Para isso, são realizadas análises das métricas de desempenho: precisão, probabilidade de detecção e taxa de acerto média.

Além desta introdução, este artigo está estruturado em quatro seções que abordam os seguintes conteúdos: A Seção 2 apresentará alguns trabalhos relacionados com o tema. A Seção 3 apresentará o método proposto para a classificação automática de áreas queimadas. A Seção 4 apresentará os resultados da aplicação do método proposto. Por fim, a Seção 5 apresentará as conclusões obtidas com base na discussão dos resultados, assim como, contribuições geradas pela pesquisa. Também são apresentadas algumas recomendações e propostas para trabalhos futuros.

2 TRABALHOS RELACIONADOS

Em [9], os autores propõem o desenvolvimento de um método para a classificação de áreas próximas, capaz de determinar polígonos queimados e não queimados no território brasileiro. A estratégia é usar uma base de conhecimento, gerada a partir de conjuntos de dados de imagens do satélite Landsat 8, de áreas queimadas e não queimadas. Os autores destacam que esta tarefa é um grande desafio devido à complexidade de desenvolver um classificador automático preciso que tenha a capacidade de detectar áreas queimadas em regiões para as quais não existem dados previamente validados. O trabalho aplica o modelo de aprendizado de máquina *Random Forest* para promover a classificação automática. A justificativa desta escolha é que este modelo apresentou bons resultados em trabalhos anteriores e possui interpretabilidade simples. Os resultados obtidos mostram que é viável implementar a estratégia de usar conjuntos de dados próximos para caracterizar e classificar locais que apresentam dados ausentes ou insuficientes, visto que, os experimentos apresentaram alta precisão de classificação.

A proposta apresentada em [4] consiste em analisar a acurácia do mapeamento de incêndios florestais em terras indígenas usando dados de satélite. O método aplica o modelo Árvore de Decisão em um conjunto de dados sobre a temperatura da superfície e o Índice de Queima Normalizada (do Inglês, *Normalized Burn Ratio*) (NBR). Os autores destacam que a abordagem baseada em limiarizações automáticas, calibradas por meio da distribuição espacial da temperatura da superfície e do índice NBR, é capaz de gerar valores altos de coeficiente de separabilidade de áreas queimadas e não queimadas. Como resultado, foi possível observar que a temperatura da superfície e o índice NBR permitem construir

um limiar adaptativo para identificação de áreas queimadas, apresentando acurácia global de 90%.

O artigo [12] descreve uma abordagem para melhorar a classificação semiautomática de áreas queimadas por meio da aplicação de Redes Neurais Artificiais (RNAs). Esta abordagem reavalia as cicatrizes e automaticamente separa as cicatrizes de queimadas, cicatrizes de áreas de colheitas e desmatamento. Para a validação dos resultados, foi realizada uma comparação entre as áreas classificadas manualmente, por especialistas, e as áreas classificadas pela abordagem proposta. O autores destacam que os resultados mostraram que o treinamento de RNAs, em um conjunto de dados previamente conhecidos, para geração de modelos de classificação de dados de queimadas apresenta-se como uma abordagem promissora.

A principal contribuição do presente trabalho é determinar se um modelo preditivo, treinado com amostras de queimadas de 2018 é capaz de caracterizar e classificar ocorrências de queimadas do ano seguinte, com confiabilidade próxima a 95%. Além disso, busca-se discutir a aplicação de uma ferramenta de AutoML, com a finalidade de apoiar a escolha de um modelo de aprendizado de máquina para essa aplicação específica.

3 DESCRIÇÃO DO MÉTODO PROPOSTO PARA A CLASSIFICAÇÃO DE ÁREAS QUEIMADAS

O método proposto neste trabalho é baseado em dois conjuntos de dados disponibilizados pelo Programa Queimadas. Os conjuntos de dados são referentes a ocorrências de queimadas. De modo geral, o método abrange as etapas do processo de descoberta de conhecimento. A Figura 1 ilustra a visão geral do método proposto para a classificação automática de áreas queimadas.

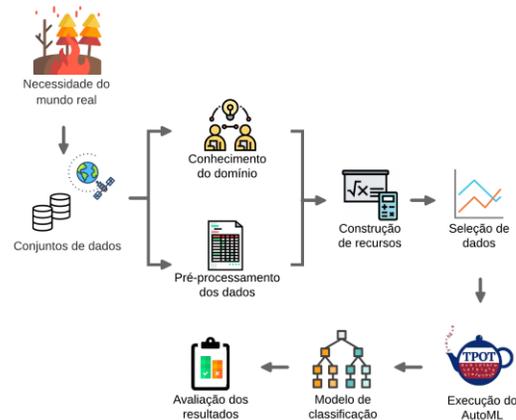


Figura 1: Visão geral do método proposto para a classificação automática de áreas queimadas.

Inicialmente, observa-se a necessidade do mundo real de automatizar as tarefas relacionadas a caracterização e a classificação de áreas queimadas. Em um segundo momento, foram adquiridos os conjuntos de dados históricas referentes a ocorrências de queimadas dos anos de 2018 e 2019.

A etapa de pré-processamento de dados ocorreu juntamente com o estudo do domínio. No início do pré-processamento dos dados foram realizadas reuniões com especialistas da área de queimadas com o objetivo de favorecer o entendimento dos dados e para a criação de estratégias.

Na sequência, foi realizada a construção de recursos. Esta etapa consiste em gerar novos atributos a partir do conjunto de dados de entrada.

A fase seguinte consiste na seleção de dados. Nesta etapa, a finalidade é selecionar atributos de interesse e obter uma representação reduzida do conjunto de dados. A partir deste ponto, considera-se que os dados foram transformados em características e estão prontos para serem aplicados em algoritmos de aprendizado supervisionado.

O próximo passo consistiu na execução do AutoML, cujo objetivo foi descobrir o modelo de aprendizado de máquina indicado pela ferramenta. Por fim, foi realizada a aplicação do modelo de classificação. A finalidade é gerar um modelo preditivo que, a partir de um novo conjunto de dados, seja capaz de evidenciar padrões e promover a classificação de áreas queimadas.

O último passo consiste na avaliação das métricas de desempenho: precisão, probabilidade de detecção e taxa de acerto média, para cada um dos experimentos construídos.

3.1 Identificação e Compreensão dos Conjuntos de Dados

Neste artigo, foram usadas cenas do sensor *Operational Land Imager* (OLI), a bordo do satélite Landsat 8. A resolução espacial deste satélite é de 30 metros e período de revisita é de 16 dias. O Landsat 8 possui uma série histórica de cenas disponíveis do território brasileiro, sendo que para este trabalho foram consideradas as cenas nas órbitas-ponto: 219/064, 219/065, 219/066, 220/064, 220/065, 220/066, 221/065 e 221/066, referentes aos anos de 2018 e 2019. A Figura 2 ilustra a área de estudo, com destaque para as órbitas-ponto do satélite Landsat 8.

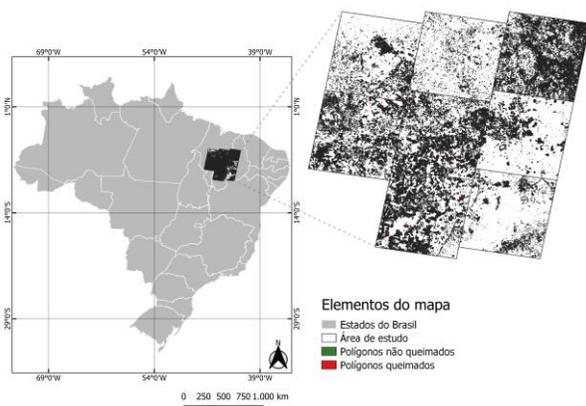


Figura 2: Recorte das cenas para as órbitas-ponto da área de estudo.

O conjunto de dados de 2018 consiste em 118881 polígonos, nos quais, 75060 são rotulados como “Não-queimados” e 43821 são rotulados como “Queimados”. Enquanto que, o conjunto de

dados de 2019 consiste em 226459 polígonos, nos quais, 149224 são rotulados como “Não-queimados” e 77235 são rotulados como “Queimados”.

De acordo com [13], após o incêndio, ocorrem uma série de mudanças de espectro em função do fogo consumir a vegetação, destruindo a clorofila, deixando o solo descoberto, carbonizando as raízes e alterando a umidade do solo. A redução da clorofila resulta em um aumento na região visível do espectro eletromagnético e na diminuição na região próxima ao infravermelho. Com a diminuição da copa das árvores e da umidade do solo, a região do infravermelho médio aumenta após um incêndio. Por este motivo, a análise de imagens Landsat é considerada uma abordagem valiosa para o mapeamento de incêndios. Dentre os procedimentos usados para classificação de áreas queimadas, destaca-se o uso de índices espectrais, derivados das bandas do Landsat, de diferenças temporais pré-incêndio e pós-incêndio.

Os conjuntos de dados são formados pelas medianas dos valores das bandas espectrais 3, 4, 5, 6 e 7, para todos os pixels do polígono identificado, medianas dos valores das bandas 3, 4, 5, 6 e 7, para todos os pixels da imagem anterior que compreendem o polígono identificado, diferenças entre os valores das medianas anteriores para as bandas espectrais 5, 6 e 7, data de passagem do satélite, identificador de órbita-ponto do Landsat 8, área do polígono (em hectares) e classe do polígono validada por um auditor.

A Figura 3 ilustra duas imagens de cobertura da Terra, pré-incêndio e pós-incêndio, do ponto com coordenadas geográficas $7^{\circ}21'36.000''S$ $42^{\circ}40'48.000''W$. Estas imagens foram obtidas por meio do portal *Brazil Data Cube*. A primeira imagem, no período de 23 de abril de 2018, representa um ponto sem queimada, enquanto que, a segunda imagem, no período de 14 de setembro de 2018, representa um ponto com cicatriz de queimada.



Figura 3: Imagens da cobertura do ponto de região ($7^{\circ}21'36.000''S$ $42^{\circ}40'48.000''W$) do satélite Landsat 8.

3.2 Pré-processamento de Dados

Na etapa de pré-processamento é necessário aplicar codificadores de etiquetas nas variáveis dos conjuntos de dados. Os dados de textos categóricos devem ser convertidos em dados numéricos para, posteriormente, serem compreendidos por modelos de predição. A variável de saída “Classe” é do tipo categórica. Logo, foi transformada para o tipo numérico por meio do uso da classe *LabelEncoder* da biblioteca *Scikit-learn* da linguagem de programação Python.

Quanto às variáveis de entrada, sendo estas representadas pelas bandas e índices espectrais, foi observado que estas variáveis são numéricas, logo, não é necessário aplicar a abordagem anterior. Porém, estas variáveis podem apresentar valores discrepantes umas

em relação às outras, isto é, as variáveis de entrada podem estar em escalas distintas. Esse fator pode causar problemas no treinamento do modelo, uma vez que, escalas diferentes podem gerar influência nos resultados finais. Para solucionar esse problema foi realizada a padronização das variáveis por meio da classe *StandardScaler*, também da biblioteca *Scikit-learn*.

3.3 Construção de Recursos

A construção de recursos consiste no processo de criação de novos atributos a partir dos dados existentes. Essa etapa é de grande importância, porque os algoritmos de aprendizado de máquina aprendem com os dados fornecidos. A finalidade é extrair recursos para construir representações que sejam capazes de classificar os conjuntos de dados. Neste trabalho, foram criados recursos a partir dos valores de bandas espectrais para se obter os índices espectrais, NDVI, BAI e NBR, explicados a seguir.

Nos trabalhos [9][13][14], foi testado com sucesso a capacidade da combinação de índices espectrais para caracterizar as áreas queimadas. Um dos mais usados, tanto do ponto de vista do pós-incêndio quanto da diferença pré-incêndio e pós-incêndio, é o Índice de Vegetação por Diferença Normalizada (do Inglês, *Normalized Difference Vegetation Index*) (NDVI). A justificativa é que são observadas quedas importantes nos valores deste índice nas áreas queimadas.

Segundo [13], o NDVI é um indicador da biomassa fotossinteticamente ativa. Este índice ajuda a diferenciar a vegetação de outros tipos de cobertura da superfície, bem como identificar e classificar áreas cultivadas em mapas temáticos, auxiliando na detecção de mudanças de padrões. O NDVI é calculado por meio da diferença entre a reflectância das bandas infravermelho próximo (NIR) e do vermelho (RED), dividida pela soma das duas reflectâncias, conforme a Equação 1.

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (1)$$

Além do NDVI, neste trabalho também foi usado o Índice de Área Queimada (do Inglês, *Burn Area Index*) (BAI). Este índice destaca a área queimada por meio da combinação do espectro do vermelho (RED) ao infravermelho próximo (NIR), enfatizando alvos carbonizados em imagens pós-incêndio [14]. O BAI é calculado por meio da Equação 2.

$$BAI = \frac{1}{(0,1 + RED)^2 + (0,06 + NIR)^2} \quad (2)$$

O Índice de Queimada Normalizada (do Inglês, *Normalized Burn Ratio*) (NBR) é usado na detecção de cicatrizes de queimadas e avaliação da severidade do fogo, uma vez que nas composições do índice são usadas as regiões do espectro eletromagnético que sofrem alterações após a queima. Em áreas com vegetação o índice assume valores positivos, enquanto que, em áreas de solo descoberto seus valores são negativos [13]. O NBR é obtido por meio da Equação 3.

$$NBR = \frac{NIR - SWIR}{NIR + SWIR} \quad (3)$$

3.4 Seleção de Dados

A fase seguinte consiste na seleção de atributos. Os atributos selecionados foram as bandas espectrais, índices espectrais e a classe de saída rotulada por um auditor. A partir deste ponto, considera-se que os dados foram transformados em características e estão prontos para serem aplicados em um modelo de classificação supervisionada.

A saída desta etapa resultou em dois conjuntos de dados que contém as medianas dos valores das bandas espectrais de todos os pixels dos polígonos identificados, medianas dos valores das bandas espectrais para todos os pixels das imagens anteriores que compreendem os polígonos identificados e as diferenças entre os valores das bandas espectrais das imagens pré-incêndio e pós-incêndio.

3.5 Aplicação de Aprendizado de Máquina Automatizado

O AutoML tem como objetivo automatizar as tarefas essenciais para a implementação de modelos de aprendizado de máquina. Os métodos baseados em AutoML automatizam uma ou mais etapas do processo de geração de um modelo como, por exemplo, construção de recursos, seleção de recursos, processamento de recursos, seleção de algoritmos de aprendizado de máquina e otimização de hiperparâmetros [10][11].

Devido à existência de uma grande variedade de modelos e suas respectivas configurações de hiperparâmetros, a tarefa de projetar e ajustar sistemas de aprendizado de máquina é intensiva. As tecnologias baseadas em AutoML estão focadas em automatizar esta tarefa. Estas tecnologias permitem aos usuários desenvolver modelos úteis para a aplicação específica de forma ágil.

Existem diversas ferramentas de AutoML como, por exemplo, TPOT, Autokeras, PennAi, H2O e AutoWeka. Neste trabalho foi usada o *framework* TPOT. A justificativa desta escolha é que o TPOT é de código aberto, está bem documentado e em desenvolvimento ativo no período de escolha de ferramentas. De acordo com [11], o *framework* tem a capacidade de avaliar *pipelines* de forma automática e eficiente. Em seu estudo [7], o autor destaca que uma das características mais importantes do TPOT é a capacidade de exportar um modelo preditivo e inseri-lo diretamente em um código-fonte.

O TPOT é baseado na linguagem de programação Python. O *framework* otimiza *pipelines* de aprendizado de máquina para produzir modelos de regressão ou de classificação. O TPOT explora vários canais possíveis para encontrar o melhor dentre eles para uma base de dados específica. Para isso, são considerados vários algoritmos de aprendizado de máquina, bem como diversas configurações de hiperparâmetros [10].

Para construir e otimizar *pipelines*, o TPOT usa programação genética. A ideia do algoritmo evolutivo é criar, inicialmente, uma população de *pipelines* de aprendizado de máquina de forma aleatória e evolui-la com mutações e cruzamentos de geração em geração [1]. Os *pipelines* são avaliados e recebem uma pontuação de condicionamento físico, para que o procedimento de seleção possa determinar quais indivíduos estarão na próxima geração.

Existem algumas restrições de parada, geralmente, estão relacionadas ao tempo máximo de execução ou tamanho da população.

Inicialmente, antes da aplicação do TPOT, é preciso realizar a preparação e limpeza dos conjuntos de dados brutos. Como algoritmos de aprendizado de máquina aprendem de diferentes maneiras, em alguns casos é necessário incluir algo além dos dados de entrada pré-processados para generalizar o conteúdo. A construção, o processamento e a seleção de recursos, podem facilitar o entendimento dos conjuntos de dados de entrada, por meio da adição de regras implícitas. Como resultado deste processo iterativo tem-se a seleção de um modelo de aprendizado de máquina. Na sequência, é realizada a otimização de hiperparâmetros que consiste em ajustar diferentes restrições, pesos ou taxas de aprendizagem para encontrar operadores que minimizem os erros de predição do modelo selecionado. Por fim, a avaliação do modelo resultante é de responsabilidade do usuário. A Figura 4 ilustra a visão geral do processo de aprendizado de máquina realizado por TPOT.

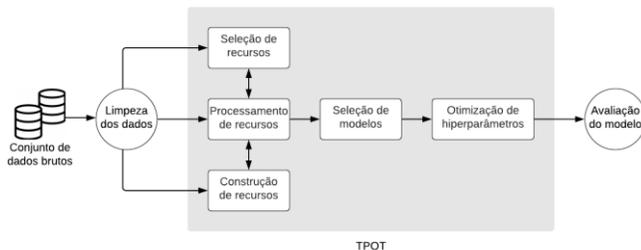


Figura 4: Visão geral do processo de aprendizado de máquina automatizado por TPOT.

Neste trabalho, inicialmente, o TPOT foi executado para descobrir o modelo de aprendizado de máquina indicado para os conjuntos de dados específicos. Em um segundo momento, ocorreu a implementação do modelo preditivo indicado pelo *framework*. As Seções 3.5.1 e 3.5.2 descrevem mais detalhes destes dois processos, respectivamente.

3.5.1 Descobrimto do Modelo de Aprendizado de Máquina.

Inicialmente, os dados são divididos em conjuntos de treinamento e teste. A porcentagem do conjunto de dados para treinamento é de 70%, enquanto que a porcentagem do conjunto de dados para teste é de 30%. Somente para o Experimento 3, o qual visa responder a pergunta de pesquisa, todo o conjunto de dados de 2018 foi usado para treinamento, enquanto que, todo o conjunto de dados de 2019 foi usado para teste.

A função *Fit* inicializa o algoritmo de programação genética para encontrar o *pipeline* com a pontuação mais alta com base na validação cruzada média de dobras K. Em seguida, o *pipeline* é treinado em todo o conjunto de amostras fornecidas e a instância TPOT pode ser usada como um modelo adequado. Na sequência, o *pipeline* pode ser avaliado no conjunto de dados de teste por meio da função *Score*. A ferramenta retorna a pontuação do *pipeline* otimizado nos dados de teste usando a função de pontuação

especificada pelo usuário. A função de pontuação padrão do TPOT, para problemas de classificação, é a precisão.

A saída desta etapa será um modelo de classificação, que pode ser usado para prever a classe de novas instâncias sem rótulo. O TPOT gerará um arquivo, em linguagem de programação Python, com o modelo de classificação resultante e suas respectivas configurações de hiperparâmetros.

3.5.2 Execução do Modelo de Aprendizado de Máquina.

Após a descoberta do modelo de classificação mais apropriado para o conjunto de dados, a próxima etapa consiste na execução do modelo de aprendizado de máquina selecionado.

O TPOT possui operadores padrão para os modelos de aprendizado de máquina. Além das configurações padrão que acompanham o TPOT, o *framework* permite que os usuários forneçam configurações personalizadas para seus operadores e parâmetros. Neste trabalho, em ambos os experimentos realizados foram usadas as configurações padrão.

Inicialmente, é necessário aplicar codificadores de etiquetas nas variáveis categóricas dos conjuntos de dados. Novamente, a variável de saída “Classe” foi transformada para o tipo numérico. Quanto às variáveis de entrada, estas passaram por uma normalização.

Na sequência, o conjunto de dados de entrada foi particionado em amostra de treinamento e teste. A porcentagem do conjunto de dados para treinamento foi de 70%, enquanto que, a porcentagem do conjunto de dados para teste foi de 30%. O *pipeline* foi treinado em todo o conjunto de dados de treinamento. Em seguida, o modelo foi avaliado no conjunto de dados de teste. Por fim, para permitir a avaliação do classificador, foram geradas as matrizes de confusão e calculadas as métricas de desempenho.

3.6 Avaliação de Padrões

Na área de aprendizado de máquina, existem diversas métricas que podem ser usadas para avaliação da qualidade de modelos de regressão e de classificação. A escolha das métricas depende do contexto do problema que se pretende avaliar. Algumas medidas mais comumente usadas para avaliar a qualidade de modelos de classificação são: taxa de acerto média, precisão, revocação e medida F. O presente artigo se utilizará destas métricas para avaliar a classificação de áreas queimadas. Os tópicos a seguir explicam os significados de cada uma dessas métricas.

- Taxa de acerto média: consiste na razão entre o número de acertos pelo número total de exemplos. Esta métrica deve ser usada em aplicações em que as classes de saídas são de mesma proporção. O uso da taxa de acerto em problemas com classes desproporcionais, pode causar uma falsa impressão de bom desempenho.
- Precisão: visa identificar a porcentagem das amostras classificadas positivamente, isto é, quantas amostras foram classificadas de forma correta em relação ao número de classificações positivas.
- Revocação: também conhecida como probabilidade de detecção, é definida pelo o número de exemplos

classificados como pertencentes a uma determinada classe, que realmente são daquela classe, dividido pela quantidade total de exemplos que pertencem a esta classe, mesmo que sejam classificados em outra.

- Medida F: é obtida por meio da média harmônica entre as métricas de avaliação, precisão e revocação. A medida F consiste em uma combinação dos valores de precisão e de revocação.

Para avaliar o desempenho dos algoritmos de classificação podem ser usados métodos de validação cruzada. Neste trabalho, utilizou-se a validação cruzada do tipo *K-fold*. De acordo com [16], a metodologia de validação cruzada *K-fold* consiste em dividir o conjunto de dados de entrada em *K* subgrupos distintos, também chamados de *folds*, com tamanhos aproximadamente iguais. O modelo de aprendizado de máquina é treinado em todos os subconjuntos, exceto um ($K - 1$). Na sequência, avalia-se o modelo no subconjunto que não foi usado para o treinamento. O processo é repetido *K* vezes, com um subconjunto diferente reservado para avaliação e excluído do treinamento. Ao término das iterações, calcula-se a média das métricas de desempenho. A Figura 5 exemplifica uma representação do método de validação cruzada *K-fold* para *K* iterações.

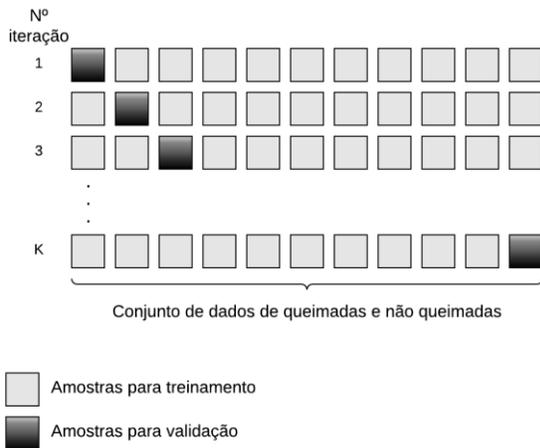


Figura 5: Representação da metodologia de validação cruzada *K-fold*.

De acordo com [15], o viés do método *K-fold* diminui quanto maior o valor de *K*. Entretanto, um valor de *K* muito elevado pode causar o aumento do custo computacional da implementação, além de implicar em uma amostra de validação muito pequena, o que aumenta a variância. Na literatura, se discute qual valor de *K* seria o ideal. As opções mais usuais são os valores $K = 2$, $K = 5$ e $K = 10$. Em [15], relata-se que $K = 10$ apresentou melhor desempenho para a aplicação desenvolvida. Neste trabalho, adotou-se o número de partições com $K = 10$.

Para avaliar o método de classificação automática de áreas queimadas foram gerados três experimentos distintos, o que os diferenciam são os conjuntos de dados de treinamento e testes usados. Em todas as situações, foram realizados os processos descritos nas Seções 3.5.1 e 3.2 com o objetivo de descobrir o modelo de aprendizado de máquina mais adequado para os dados

de entrada e os resultados de sua aplicação, respectivamente. Em especial, foram construídos três experimentos para avaliar a qualidade do classificador. A descrição dos experimentos é apresentada nos tópicos a seguir.

- Experimento 1: O conjunto de dados de 2018 foi usado para treinamento e teste na proporção de 70% e 30%, respectivamente.
- Experimento 2: O conjunto de dados de 2019 foi usado para treinamento e teste na proporção de 70% e 30%, respectivamente.
- Experimento 3: O conjunto de dados de 2018 foi integralmente usado para treinamento, enquanto que, o conjunto de dados de 2019 foi usado para teste. Este experimento visa responder a pergunta de pesquisa.

4 AVALIAÇÃO DOS RESULTADOS OBTIDOS COM O MÉTODO PROPOSTO

4.1 Avaliação do Experimento 1

No primeiro experimento elaborado neste trabalho, a construção do conjunto de atributos de entrada foi baseada em recursos de bandas e índices espectrais dos dados de queimadas do ano de 2018. Nesse contexto, buscou-se avaliar o desempenho de um modelo de aprendizado de máquina na classificação de amostras do mesmo ano. As Seções 4.1.1 e 4.1.2 descrevem o modelo selecionado pelo AutoML e os resultados obtidos por este modelo, respectivamente.

4.1.1 Execução do Experimento 1.

Para este experimento, foi executado o otimizador de *pipelines* da biblioteca TPOT. A execução do AutoML resultou no modelo de classificação *K-Neighbors*. O tempo de execução do *framework* para indicar esse modelo foi de 3h 17min 31s. O Algoritmo 1 apresenta o modelo selecionado pelo AutoML com suas respectivas configurações de hiperparâmetros.

```
exported_pipeline = make_pipeline(
    RobustScaler(),
    KNeighborsClassifier(n_neighbors=50, p=2, weights="distance")
)
```

Algoritmo 1: Modelo selecionado para o Experimento 1.

4.1.2 Análise do Experimento 1.

A Tabela 1 apresenta a matriz de confusão para este experimento. A taxa de acerto média foi de 98,13% com desvio padrão de 0,17%.

Tabela 1: Matriz de Confusão para o Experimento 1.

		Classe predita	
		Queimada	Não queimada
Classe real	Queimada	12886	264
	Não-queimada	402	22113

A Tabela 2 apresenta os resultados das métricas de desempenho do classificador para o Experimento 1. É possível destacar que a medida de Precisão e de Revocação (também conhecida como probabilidade de detecção), para as duas classes, apresentaram pequenas diferenças. Esse fato sugere que o preditor não é tendencioso para apenas uma das classes de saída. Outra observação importante é que todos os valores das métricas de desempenho foram superiores a 95%, que é o valor desejável para considerar um método automático de classificação de áreas queimadas.

Tabela 2: Métricas de Avaliação para o Experimento 1.

Classe	Métricas de avaliação		
	Precisão	Revocação	Medida F
Queimada	0,97	0,98	0,97
Não-queimada	0,99	0,98	0,99

4.2 Avaliação do Experimento 2

No segundo experimento, considerou-se os mesmos atributos de entrada do experimento anterior. Porém, desta vez, foram usados as amostras de ocorrências de queimadas do ano de 2019. As Seções 4.2.1 e 4.2.2 descrevem o modelo selecionado pelo AutoML e os resultados obtidos por este modelo, respectivamente.

4.2.1 Execução do Experimento 2.

A execução do AutoML resultou no modelo de classificação *Decision Tree*. O tempo de execução do *framework* foi de 3h 53min 9s. O Algoritmo 2 apresenta o modelo selecionado pelo AutoML com suas respectivas configurações de hiperparâmetros.

```
exported_pipeline = make_pipeline(
    StackingEstimator(estimator=MultinomialNB(alpha=100.0,
                                              fit_prior=False)),
    DecisionTreeClassifier(criterion="entropy",
                          max_depth=10,
                          min_samples_leaf=14,
                          min_samples_split=10)
)
```

Algoritmo 2: Modelo selecionado para o Experimento 2.

4.2.2 Análise do Experimento 2.

Para este experimento, a taxa de acerto média foi de 96,01% com desvio padrão de 0,23%. A Tabela 3 apresenta a matriz de confusão.

Tabela 3: Matriz de Confusão para o Experimento 2.

Classe real	Classe predita	Classe predita	
		Queimada	Não queimada
Queimada	Queimada	21808	1369
Não-queimada	Queimada	1352	43409

Pode-se observar que o conjunto de dados de 2019 possui uma maior quantidade de amostras do que o conjunto de dados de 2018.

Para facilitar a comparação é gerada a Tabela 4 que representa as métricas de desempenho deste experimento. Pode-se observar que as métricas, para a classe “Queimada”, foram inferiores aos resultados obtidos para a classe “Não-queimada”. Porém, todas as métricas foram muito próximas a 95%.

Tabela 4: Métricas de Avaliação para o Experimento 2.

Classe	Métricas de avaliação		
	Precisão	Revocação	Medida F
Queimada	0,94	0,94	0,94
Não-queimada	0,97	0,97	0,97

4.3 Avaliação do Experimento 3

Neste experimento foram considerados os recursos espectrais dos anos de 2018 e 2019. A razão pela qual foi construída a junção dos anos é para avaliar se um modelo de aprendizado de máquina treinado com amostras de queimadas de 2018 é capaz de classificar amostras do ano seguinte. Este experimento visa responder a pergunta de pesquisa considerada no início do trabalho. As Seções 4.3.1 e 4.3.2 descrevem o modelo selecionado pelo AutoML e os resultados obtidos por este modelo, respectivamente.

4.3.1 Execução do Experimento 3.

A execução do AutoML resultou no modelo *Decision Tree*. O tempo de execução do *framework* foi de 4h 27min 5s. O Algoritmo 3 apresenta o modelo selecionado pelo AutoML com suas respectivas configurações de hiperparâmetros.

```
exported_pipeline = make_pipeline(
    StackingEstimator(estimator=MultinomialNB(alpha=100.0,
                                              fit_prior=False)),
    DecisionTreeClassifier(criterion="entropy",
                          max_depth=10,
                          min_samples_leaf=14,
                          min_samples_split=10)
)
```

Algoritmo 3: Modelo selecionado para o Experimento 3.

4.3.2 Análise do Experimento 3.

A taxa de acerto média foi de 95,36% com desvio padrão de 2,16%. A Tabela 5 apresenta a matriz de confusão e permite visualizar a dispersão dos valores previstos para as classes reais.

Tabela 5: Matriz de Confusão para o Experimento 3.

Classe real	Classe predita	Classe predita	
		Queimada	Não queimada
Queimada	Queimada	72202	5033
Não-queimada	Queimada	5472	143752

A Tabela 6 apresenta os resultados das métricas de desempenho do Experimento 3. Pode-se observar que o modelo apresenta uma maior dificuldade na determinação da classe “Queimada”. Embora os resultados para a classe “Queimada” tenham sido inferiores aos

obtidos para a classe “Não-queimada”, as métricas de precisão, revocação e Medida F, foram muito próximas ao valor desejado de 95%. Estas evidências podem significar que o modelo é promissor, para promover a classificação automática de áreas queimadas.

Tabela 6: Métricas de Desempenho para o Experimento 3.

Classe	Métricas de avaliação		
	Precisão	Revocação	Medida F
Queimada	0,93	0,93	0,93
Não-queimada	0,97	0,96	0,96

5 CONCLUSÕES E PROPOSTAS PARA TRABALHOS FUTUROS

Os sistemas de classificação automática de áreas queimadas têm um enorme potencial de promover a redução do tempo necessário para realizar a identificação de áreas queimadas anualmente. No contexto econômico, a implantação de sistemas de classificação automática traz a expectativa de converter os trabalhos de auditoria em tratamentos de eventos duvidosos gerados por um classificador e, assim, minimizar custos e reduzir o tempo necessário para a classificação.

Este trabalho avaliou um método para promover a classificação automática de áreas queimadas para lidar com um extenso território e a complexidade de obtenção de amostras de queimadas validadas por auditores. Pode-se observar que cada experimento sugere um modelo de aprendizado de máquina diferente, não é estabelecido um só modelo e por isso é interessante usar o AutoML para sugerir boas alternativas e avaliar o melhor modelo para um conjunto de dados de entrada específico.

Analisando os resultados obtidos pelo Experimento 3, baseado no uso de amostras de 2018 e 2019, sugere-se que este método é apropriado para o desenvolvimento de um sistema de classificação automática de áreas queimadas para o ano seguinte. Uma hipótese para esta conclusão é que os valores das métricas de desempenho, obtidas pelo modelo, foram muito próximos ao valor estabelecido como precisão ótima de 95%. Estes resultados mostram que é viável usar a estratégia para promover a classificação automática de áreas queimadas.

Por fim, destaca-se que existe uma grande variedade de possibilidades para caracterizar e classificar áreas queimadas. Durante o desenvolvimento deste trabalho foram identificados alguns aspectos importantes que podem ser considerados em trabalhos futuros. Este trabalho focou no desenvolvimento e avaliação de um método de classificação supervisionada. Entretanto, é possível usar outras abordagens, bem como aplicar outras técnicas. Futuras investigações poderão identificar quais são os índices espectrais mais adequados para promover a classificação de áreas queimadas. Sabe-se que diferentes tipos de vegetação, efeitos sazonais e outras características locais podem enriquecer a classificação automática de áreas queimadas. Uma sugestão é incorporar novos recursos para treinamento de modelos de aprendizado de máquina.

AGRADECIMENTOS

Os autores agradecem ao Programa Queimadas do Instituto Nacional de Pesquisas Espaciais - INPE. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), processos 88882.461660 e 88887.603912.

REFERÊNCIAS

- [1] Adithya Balaji e Alexander Allen. 2018. Benchmarking automatic machine learning frameworks.
- [2] Instituto Nacional de Pesquisas Espaciais (INPE), 2021. Portal do Monitoramento de Queimadas e Incêndios. Disponível em: <https://queimadas.dgi.inpe.br/queimadas/portal>. Acessando em: 25 maio de 2021.
- [3] John R. Jensen. 2009. Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestres. Ed. Parêntese.
- [4] Juarez Antônio da Silva Júnior e Vanessa Silva de Moraes. 2020. Análise de Acurácia para o Monitoramento de Incêndios Florestais em Terras Indígenas Brasileiras Utilizando Classificação por Árvore de Decisão Baseado em Imagens OLI 0,85 – 0,11 μm . In *Revista Brasileira de Sensoriamento Remoto*, v. 1, n. 3, 12-28.
- [5] Julia Abrantes Rodrigues, Renata Libonati, Leonardo de Faria Peres e Alberto Setzer. 2018. Mapeamento de áreas queimadas em Unidades de Conservação da região serrana do Rio de Janeiro utilizando o satélite Landsat-8 durante a seca de 2014. In *Anuário do Instituto de Geociências – UFRJ*, v. 41, n. 1, 318-327. DOI: http://dx.doi.org/10.11137/2018_1_318_327.
- [6] Luana de Aratijo Nogueira Santiago e Rogério Santiago Lopes. 2021. Impacts on human health due to the emission of aerosols caused by burns. In *Brazilian Journal of Development*, v. 7, n. 1, 9069-9075. DOI: 10.34117/bjdv7n1-616.
- [7] Matthias Feurer, Aaron Klein, Katharina Eggensperger, Jost Springenberg, Manuel Blum e Frank Hutter. 2015. Efficient and robust automated machine learning. In *Advances in Neural Information Processing*.
- [8] Olga Oliveira Bittencourt, Fabiano Morelli, Cicero Alves dos Santos Júnior e Rafael Santos. 2019. Evaluating Classification Models in a Burned Areas' Detection Approach. In *International Conference on Computational Science and Its Applications*. 577-591. DOI: https://doi.org/10.1007/978-3-030-24305-0_43.
- [9] Olga Oliveira Bittencourt, Fabiano Morelli, Cicero Alves dos Santos Júnior e Rafael Santos. 2020. An Approach to Classify Burned Areas Using Few Previously Validated Samples. In *International Conference on Computational Science and Its Applications*. 577-591. DOI: https://doi.org/10.1007/978-3-030-58814-4_17.
- [10] Randal S. Olson, Ryan J. Urbanowicz, Peter C. Andrews, Nicole A. Lavender, La Creis Kidd e Jason H. Moore. 2016. Automating Biomedical Data Science Through Tree-Based Pipeline Optimization. In *Applications of Evolutionary Computation*, Springer, 123-137.
- [11] Randal S. Olson e Jason H. Moore. 2019. TPOT: A Tree-Based Pipeline Optimization Tool for Automating Machine Learning: Methods, Systems and Challenges. In *Automated Machine Learning*. 151-160. DOI: https://doi.org/10.1007/978-3-030-05318-5_8.
- [12] Ronaldo Nelis de Andrade, Olga Bittencourt, Fabiano Morelli, Rafael Santos. 2017. Classificação Semiautomática de Áreas Queimadas com uso de Redes Neurais. In *XVIII Brazilian Symposium on Geoinformatics*. 92-97.
- [13] S. Escuin, R. Navarro e P. Fernández. 2008. Fire severity assessment by using NBR (Normalized Burn Ratio) and NDVI (Normalized Difference Vegetation Index) derived from LANDSAT TM/ETM images. In *International Journal of Remote Sensing*, v. 29, n. 4, 1053-1073. DOI: <https://doi.org/10.1080/01431160701281072>.
- [14] S. Veraverbeke and S. Harris and S. Hook. 2011. Evaluating spectral indices for burned area discrimination using MODIS/ASTER (MASTER) airborne simulator data. In *Remote Sensing of Environment*, v. 115, 2702-2709. DOI: <https://doi.org/10.1016/j.rse.2011.06.010>.
- [15] Simone Borra e Agostino Di Ciaccio. 2010. Measuring the prediction error. A comparison of cross-validation, bootstrap and covariance penalty methods. In *Computational Statistics & Data Analysis*, Elsevier, v. 54, n. 12, 2976-2989.
- [16] Stephen Marsland. 2015. Machine learning: an algorithmic perspective. Ed. CRC Press, 2nd.
- [17] Thiago Oliveira dos Santos, Valdir Soares De Andrade Filho, Vinícius Machado Rocha e Janaína de Souza Menezes. 2017. Os impactos do desmatamento e queimadas de origem antrópica sobre o clima da Amazônia brasileira: um estudo de revisão. In *Revista Geográfica Acadêmica*, v. 11, n. 2, 157-181.