

Avaliação de técnicas de IA para auxiliar na previsão de incidência de desmatamento na Amazônia.

Bruno J. T. Fernandes¹

 orcid.org/0000-0002-6001-3925

José Roberto Osses²

 orcid.org/0000-0001-5727-2427

Rafael Façanha²

 orcid.org/0000-0003-4810-5111

¹Escola Escola Politécnica de Pernambuco, Universidade de Pernambuco, Recife, Brasil.
E-mail: bruno.fernandes@poli.br; jro2@ecom.poli.br; rfrf@ecom.poli.br

²Fundação para Inovações Tecnológicas, FITec

DOI: 10.25286/repa.v7i2.2218

Esta obra apresenta Licença Creative Commons Atribuição-Não Comercial 4.0 Internacional.

Como citar este artigo pela NBR 6023/2018: Bruno J. T. Fernandes; José Roberto Osses; Rafael Façanha Ricchezza de Farias; Avaliação de técnicas de IA para auxiliar na previsão de incidência de desmatamento na Amazonia. Revista de Engenharia e Pesquisa Aplicada, v.7, n. 2, p. 57-64, 2022.

RESUMO

Esse documento descreve as atividades de pesquisa realizados como trabalho final da Especialização em Inteligência Artificial CE Fitec 2020.1. Neste trabalho foram avaliadas aplicações de técnicas de IA para auxiliar na previsão de incidência de desmatamentos na Amazonia, baseados em dados históricos e características geográficas das áreas de estudo. Estudos apontam que 94,9% de todo o desmatamento na Amazônia brasileira ocorre dentro de 5,5 km de algum tipo de estrada ou 1,0 km de rios navegáveis. Sendo assim, 3 variáveis foram selecionadas como dados de entrada para os modelos avaliados: áreas desflorestadas em anos anteriores, distância de cursos de água e distância de rodovias. A avaliação do classificador foi feita através do processamento de imagens subsequentes da série histórica, comparando o resultado obtido pelo classificador com resultados já conhecidos.

PALAVRAS-CHAVE: Desmatamento; Amazônia; IA; Série Histórica;

ABSTRACT

This document describes the research activities carried out as the final project of the Specialization in Artificial Intelligence CE Fitec 2020.1. In this work were evaluated the application of AI techniques to help predict the incidence of deforestation in the Amazon, based on historical data and geographic characteristics of the study areas. Studies show that 94.9% of all deforestation in the Brazilian Amazon occurs within 5.5 km of some type of road or 1.0 km of navigable rivers. Thus, 3 variables were selected as input data for the evaluated models: deforested areas in previous years, distance from watercourses and distance from roads. The evaluation of the classifiers was carried out through the processing of subsequent images of the historical series, comparing the result obtained by the classifier with already known results.

KEY-WORDS: Deforestation; Brazilian Amazon; Time Series;

1 INTRODUÇÃO

Há milhares de anos a humanidade iniciou um processo de transição de um estilo de vida de caçadores-coletores para o cultivo e criação de animais para alimentação [2]. Esta transição iniciou uma dinâmica de alteração das características naturais das florestas existentes por motivos tais como: preparar a terra para a produção agrícola, criação de áreas de pastagem para gados e outros animais, extração de madeira, fibra, construção de estradas ou mineração. Este processo se acelerou e continua até hoje causando alterações climáticas que a cada ano se intensificam e afetam diretamente a humanidade.

Alguns impactos são percebidos na saúde das pessoas devido à degradação de necessidades básicas, tais como qualidade do ar, água potável, alimentos suficientes e segurança de habitação, que podem ser afetados por enchentes, incêndios entre outros fenômenos climáticos. Outros impactos estão relacionados a questões econômicas, tais como o aumento de preços sobre itens de subsistência e desemprego.

Diversas iniciativas governamentais e da sociedade em geral tem crescido no intuito de minimizar os impactos da destruição dos ecossistemas e a piora no quadro das mudanças climáticas. A conscientização da sociedade em geral, no sentido de diminuir os impactos, encontra um grande obstáculo no ganho econômico que é obtido com as atividades predadoras da natureza.

Para prevenir atividades de desflorestamento é necessário conhecer sua dinâmica e os fatores que a influenciaram historicamente [3]. Para lidar com previsões em grandes bases de dados, as técnicas de Inteligência Artificial (IA) são as mais adequadas e se tornam um caminho lógico a ser seguido.

Neste trabalho serão investigadas técnicas de Inteligência Artificial para analisar e prever a distribuição espacial do desflorestamento na Amazonia, usando a série histórica de Imagens de Satélites coletadas pelo INPE desde 1975.

A análise da série histórica fornecerá entradas para a modelagem, que tem como objetivo determinar a tendência e a probabilidade de cada localização, que atualmente contenha floresta, ser desflorestada no próximo período.

Neste trabalho, cada entrada para o algoritmo terá sua localização espacial, gerando então diversas camadas que serão sobrepostas às imagens de satélite.

Devido ao tamanho da Amazônia, será selecionada uma região com tamanho adequado a nossa capacidade de processamento, e esta será utilizada como área de estudo.

A imagem desta área no período subsequente será utilizada para comparação com o resultado probabilístico da previsão, plotado em mapa de calor, e validação da modelagem.

2 METODOLOGIA

Algoritmos de IA tem sido utilizados amplamente em processamento de imagens em diversas áreas tais como: visão computacional, carros autônomos e diagnósticos médicos.

Na área de geoprocessamento são utilizados modelos de segmentação para extrair características de imagens, resultando numa classificação automática das mesmas gerando regiões homogêneas com um conjunto de características semelhantes.

No Brasil o processo de classificação de imagens, para efeito de monitoramento do desmatamento da Amazônia, é realizado pelo INPE dentro do projeto PRODES.

Neste trabalho serão utilizados os dados históricos gerados pelo projeto PRODES como dados de entrada para modelos de IA, com o objetivo de auxiliar na previsão de incidência de desmatamento na Amazônia.

A tendência de desflorestamento em uma determinada localidade é influenciada por fatores geográficos, ambientais e socioeconômicos. A partir da seleção de um conjunto de variáveis, modelos de IA podem ser treinados com dados históricos permitindo que, a partir do modelo, sejam calculadas probabilidades de um desflorestamento ocorrer futuramente em uma região.

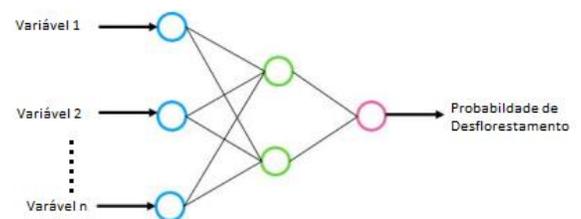


Figura 1 – Treinamento de Modelos para cálculo de probabilidade de desflorestamento.

A partir das probabilidades calculadas novas imagens podem ser geradas a utilizando a taxa de desmatamento estimada para o próximo período.

Os dados de desflorestamento gerados pelo INPE são disponibilizados em formato de planilhas, para as informações tabulares, e formato GeoTiff [6], para os dados geográficos. A manipulação dos dados geográficos é feita utilizando a biblioteca GDAL [7]. Para a manipulação dos dados tabulares são utilizadas as bibliotecas, Numpy e Pandas. Os classificadores utilizados são da biblioteca scikit-learn [8].

2 ESCOPO DO TRABALHO

2.1 CONJUNTO DE DADOS

Neste trabalho serão utilizadas 2 categorias de dados:

- Dados Históricos de desflorestamento coletados, analisados e disponibilizados pelo [INPE](#) dentro do projeto PRODES [4].
- Dados Fixos ou pouco mutáveis, tais como mapas da base cartográfica do IBGE, ou mapas gerados por órgãos governamentais, por exemplo, INCRA, FUNAI etc. Desta categoria podemos citar: Rodovias, Drenagem, Topografia, Tipo do Solo, Tipo de Vegetação, Área Indígena, Área de Proteção ambiental entre outros.

2.2 SELEÇÃO DE VARIÁVEIS DE ENTRADA

A ocorrência de um desmatamento por ação humana, para ser viável economicamente, necessita que exista um caminho até a floresta. Caso seja uma atividade extrativista se faz necessário também uma forma de transportar a madeira, o que normalmente ocorre por rodovias ou cursos de água. Existem estudos que apontam que 94,9% de todo o desmatamento na Amazônia brasileira ocorre dentro de 5,5 km de algum tipo de estrada ou 1,0 km de rios navegáveis [1]. Desta forma 3 variáveis são fortes candidatas a estarem presentes no modelo: intensidade de desmatamento anteriores; distância de cursos de água; distância de rodovias [5]. Outras variáveis podem ser incorporadas ao modelo conforme forem determinadas suas relevâncias durante o treinamento.

2.3 INTENSIDADE DE DESMATAMENTO

O processo de desmatamento pode ser separado em 2 componentes: ocorrência e intensidade. Uma vez que a maior parte do desmatamento de curto prazo ocorre nas proximidades de áreas desmatadas, optamos por focar na modelagem da intensidade do desmatamento.

O desmatamento foi então medido como a proporção de floresta substituída por ação antrópica, em uma área específica, durante um determinado intervalo de tempo, quantificado como a mudança proporcional no número de pixels da floresta com relação ao período anterior.

A variável de intensidade é a proporção de pixels desmatados dentro de uma máscara centralizada no pixel de interesse.

A Figura 2 ilustra o pré-processamento realizado nas imagens para obter os valores de intensidade para cada pixel. O tamanho da máscara utilizada vai indicar a que distância uma área desmatada irá influenciar no desflorestamento no próximo período. Nesse trabalho estamos utilizando pixels com resolução de 60x60 metros.

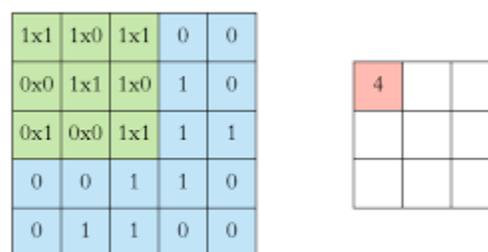


Figura 2 - Pré-processamento de intensidade de desmatamento.

Com a resolução de 60 metros estamos utilizando uma máscara com 20x20 pixels abrangendo uma área de 1.200x1.200 metros. O aumento no tamanho da máscara implica numa maior dispersão das intensidades causando uma diminuição da acurácia no modelo proposto.

3 TREINAMENTO E SELEÇÃO DE CLASSIFICADORES

Para a área de estudo selecionada foram gerados dados de diversos períodos diferentes. Foram treinados diversos classificadores permitindo assim

analisar a acurácia para cada um deles e selecionar o melhor a ser aplicado.

O primeiro levantamento de desflorestamento na Amazônia ocorreu no ano de 1975. Para este período não é possível conseguir uma boa acurácia para qualquer classificador devido ao não balanceamento da distribuição de classes dentro da área. Como o desflorestamento anterior é uma variável de entrada para o classificador, a ausência dessa informação prejudica a convergência no treinamento.

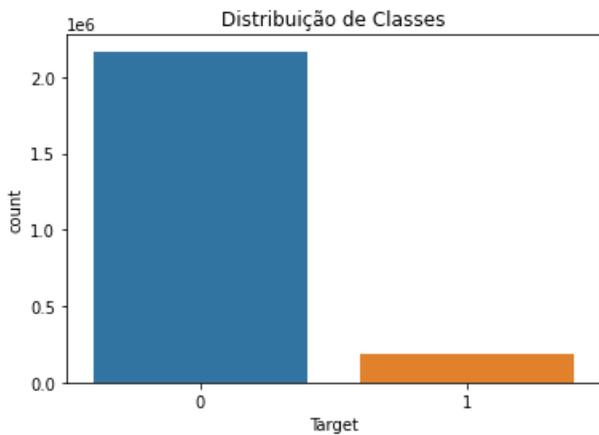


Figura 3 – Distribuição de classes para os dados de entrada.

A partir do segundo levantamento, que ocorreu em 1978, começa a ser criada uma série histórica que permite utilizar o desmatamento de anos anteriores como uma variável de entrada para o algoritmo. Além disso, passa a ser possível criar conjuntos de dados balanceados com um número adequado de amostras para cada classe, conforme Figura 4.

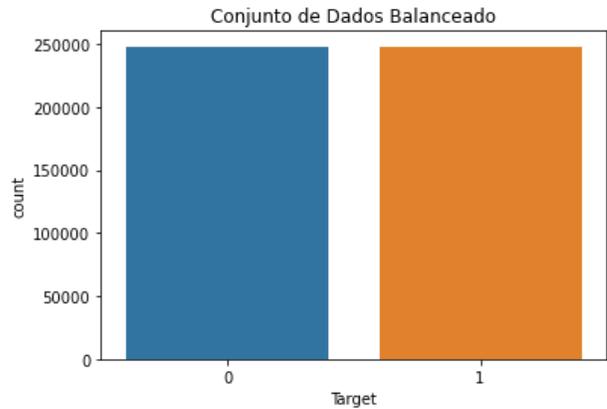


Figura 4 – Distribuição de classes para os dados após balanceamento.

Com o aumento das amostras na série histórica, os classificadores vão apresentando uma acurácia maior devido a influência das áreas previamente desmatadas.

Após o treinamento e avaliação de um conjunto de classificadores diversas configurações de áreas de estudo é possível verificar qual apresenta em média os melhores resultados. De maneira geral o “Random Forest” apresentou um resultado mais adequado em média nas diversas regiões. A Tabela 1 apresenta um exemplo de resultado de treinamento e avaliação de classificadores.

Tabela 1 – Resultado de uma avaliação de classificadores.

	Acurácia	Precisão	Recall	Score F1
Classificadores				
Regressão Logística	0.666389	0.570645	0.570645	0.588674
Linear SVC	0.742045	0.770589	0.770589	0.756728
Decision Tree	0.696061	0.670847	0.670847	0.728340
Random Forest	0.716426	0.701190	0.701190	0.746848
Gaussian NB	0.668686	0.803729	0.803729	0.651371
MLP	0.652167	0.705801	0.705801	0.663549
Adaptativo	0.686470	0.698875	0.698875	0.728943
Quadrático	0.526472	0.636928	0.636928	0.460334

4 METODOLOGIA DE CLASSIFICAÇÃO

A metodologia de classificação estabelece um processo que deve ser realizado para cada área de estudo e cada período conforme etapas abaixo:

1. Coleta de Dados

- Nesta etapa para cada pixel da imagem é aplicada a máscara com o tamanho da grade definido, sobre os dados da série histórica, obtendo o valor de intensidade de desflorestamento. As variáveis de distância de rodovias e distância de cursos d'água são obtidos com o cruzamento de informações com bases fixas de mapas de rodovias e hidrografia do IBGE. Os valores obtidos para cada pixel são armazenados em arquivo "csv" em conjunto com a posição relativa do pixel na imagem. O valor de "Target" é definido pelo estado atual do pixel em relação ao período anterior sendo: 0 para floresta no período atual e 1 para desflorestamento.

2. Treinamento do Modelo

- Os dados são divididos em treinamento e teste. O treinamento é realizado com o classificador selecionado.

3. Geração de Lista de Probabilidades

- O modelo é executado para todo o conjunto de dados gerando a lista de probabilidades para cada pixel com o método "predict_proba".
- As identificações de posicionamento e probabilidade para cada pixel são exportadas em ordem decrescente de probabilidade.

4. Seleção de Melhores candidatos.

- A partir da lista de pixels com suas respectivas probabilidades é possível selecionar os melhores candidatos de acordo com a área estimada para o desmatamento.



Figura 5 – Diagrama do processo de classificação.

5 ANÁLISE DE RESULTADOS

A região denominada Amazônia Legal é constituída de 9 estados: Acre, Amapá, Amazonas, Maranhão, Mato Grosso, Pará, Rondônia, Roraima e Tocantins.



Figura 6 – Limites geográficos da Amazonia Legal.

Neste trabalho são utilizados dados de desflorestamento obtidos e processados a partir de imagens do satélite Landsat TM (Thematic Mapper) para o período de 1975 a 2017. As imagens são subdivididas em cenas identificadas por Órbita/Ponto variando de 001/57 a 223/68.

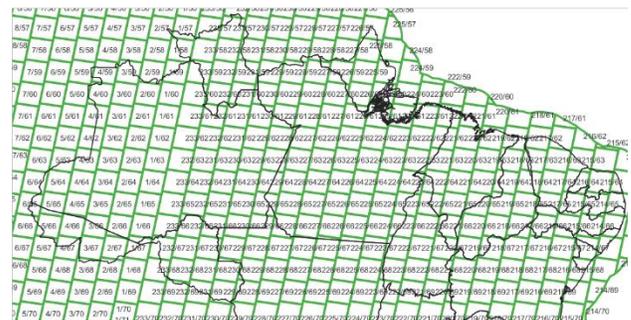


Figura 7 – Cobertura de imagens do Landsat TM para a Amazonia Legal.

A seguir são apresentados resultados de classificações utilizando IA para algumas cenas em que existem dados históricos de desmatamento.

5.1 CENA 002/67

A cena 002/67 está localizada no estado do Acre na divisa do Brasil com a Bolívia. A Figura 8 apresenta a localização geográfica de cena.

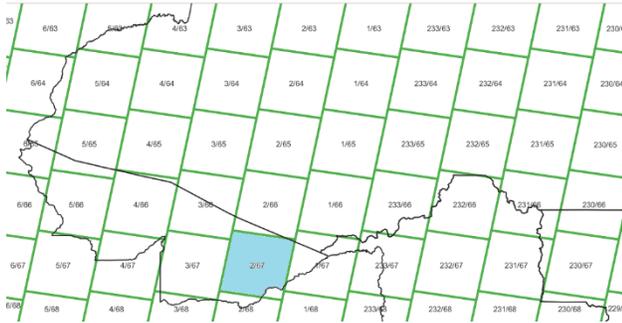


Figura 8 – Localização da cena 002/67.

A Figura 9 apresenta a situação da cena 002/67 original e atual. Na situação atual pode-se observar os desmatamentos nas proximidades de rodovias, que permitiam o acesso e escoamento da madeira

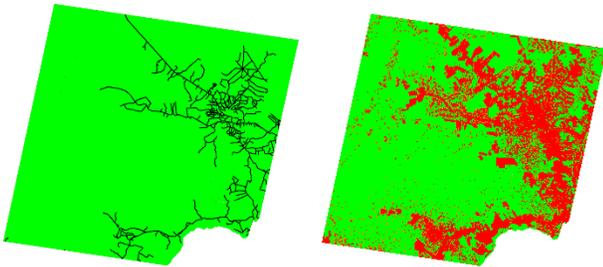


Figura 9 – Situação original da cena 002/67.

Após a obtenção de dados históricos e a criação de uma série histórica é possível executar o processo de classificação, coletando as variáveis de intensidade de desmatamento, distância de rodovias e distância de cursos d'água, e realizar o processo conforme descrito no Item 4.

A Figura 10 apresenta a situação real e estimada para a cena 002/67 usando o processo de classificação.

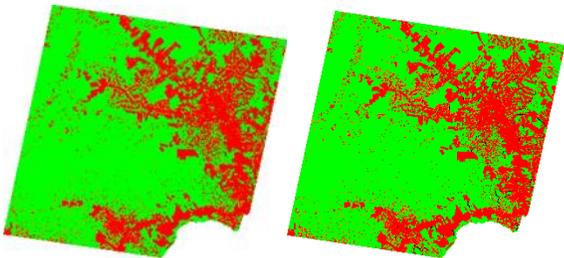


Figura 10 – Situações reais e estimadas para cena 002/67.

Para uma interpretação mais adequada do resultado é necessária uma visualização mais próxima, comparando as imagens reais e estimadas na mesma região. A sequência de Figuras de 11 a 13 apresentam 3 aproximações na visualização das imagens reais e estimadas permitindo interpretar melhor o resultado.

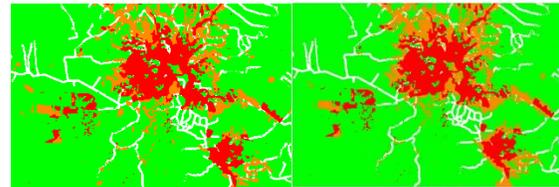


Figura 11 – Primeira aproximação da visualização para a situação para cena 002/67.

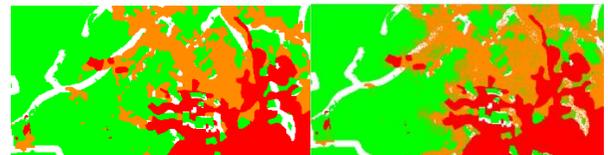


Figura 12 – Segunda aproximação da visualização para a situação estimada para cena 002/67.



Figura 13 – Terceira aproximação da visualização para a situação estimada para cena 002/67.

Com o aumento da aproximação é possível verificar que existe uma dispersão na distribuição dos pixels selecionados. Isso ocorre devido ao fato de existirem diversos locais onde as probabilidades de desmatamento calculadas pelo modelo coincidem. Nestes casos, a seleção dos pixels para estimativa terá um grau maior de incerteza.

2.2 CENA 231/67

A cena 231/67 está localizada no estado de Rondônia. Nesta região foram criados projetos de ocupação pelo Governo Federal a partir de 1960. A ocupação se iniciou no entorno da rodovia BR-364 e foram se expandindo, por iniciativas do Governo ou dos próprios camponeses, invadindo terras e demarcando lotes. A partir da ocupação, os camponeses converteram as florestas em áreas de produção agropecuária para subsistência ou comércio. A Figura 14 apresenta a localização geográfica de cena.



Figura 14 – Localização geográfica da cena 231/67.

A Figura 15 apresenta a situação real e estimada para a cena 231/67. Nesta cena pode-se observar que devido à grande proporção de área desmatada, a diferença entre a situação real e prevista é bem pequena. Isso se deve ao fato de existirem poucos pixels candidatos ao desflorestamento para este período.

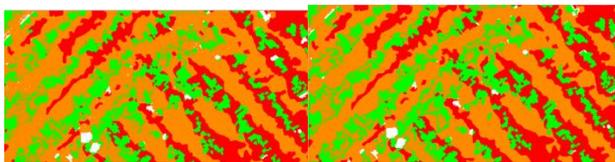


Figura 15 – Situação real e estimada para a cena 231/67.

6 CONCLUSÕES E TRABALHOS FUTUROS

Neste trabalho usamos 3 variáveis para treinamento dos classificadores: intensidade de desmatamentos anteriores, distâncias de rodovias e distâncias de cursos d'água. Conforme Jaffé et al [5], existe uma associação significativa entre novos desmatamentos e estas 3 variáveis, sendo a intensidade de desmatamentos anteriores a de maior impacto. Neste trabalho pode-se observar que em áreas com maior grau de desmatamento prévio o classificador obtém a maior acurácia. Outro fator observado de grande impacto na acurácia do treinamento do classificador, é o tamanho da máscara utilizada para e medida de intensidade do desmatamento. O tamanho da máscara tem relação direta com a distância em que o desmatamento anterior pode influenciar um novo desmatamento. Quanto maior essa distância menor será a acurácia, pois aumenta o número de pixels com a mesma intensidade de desmatamento anterior, gerando probabilidades muito próximas como saída do classificador. Isso leva o processo de seleção de melhores candidatos a desflorestamento a ter uma maior incerteza.

Neste trabalho foram utilizados dados históricos em intervalos anuais. Contudo, o processo de desmatando ocorre continuamente no decorrer do tempo e pode ser medido e modelado em períodos mais curtos. Como o desmatamento anterior tem uma grande relevância, para aprimorar o processo pode ser implementada a dinâmica de retreino dos classificadores em períodos mais curtos de forma iterativa, obtendo um novo conjunto de dados de entrada e saída para cada iteração. Desta forma o cálculo de probabilidades será realimentado com maior frequência proporcionando uma maior acurácia para seleção de pixels candidatos ao desmatamento no próximo período.

REFERÊNCIAS

- [1] Barber, C. P., Cochrane, M. A., Souza Jr, C. M., and Laurance, W. F. **Roads, deforestation, and the mitigating effect of protected areas in the amazon.** Biological conservation, 177:203–209, 2014.
- [2] GEBAUER, A. B., PRICE, T. D. **Foragers to Farmers: An Introduction. Transitions to Agriculture in Prehistory.** Madison, Prehistory Press, p. 1-10, 1992.
- [3] Alessandro Reinaldo Zabotto. **Estudos sobre impactos ambientais: Uma abordagem contemporânea. FEPAF.** Disponível em: <http://www.fepaf.org>. ISBN 978-85-7170-009-3. Acesso em: 10 ago. 2021.
- [4] Monitoramento do Desmatamento da Floresta Amazônica Brasileira por Satélite. **PROJETO PRODES.** Disponível em: <http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes>. Acesso em: 01 jul. 2021.
- [5] FORECASTING DEFORESTATION IN THE BRAZILIAN AMAZON TO PRIORITIZE CONSERVATION EFFORTS. **ENVIRONMENTAL RESEARCH LETTERS.** Disponível em: <https://iopscience.iop.org/article/10.1088/1748-9326/ac146a>. Acesso em: 12 set. 2021.
- [6] GDAL - Biblioteca de tradução para formatos de dados geoespaciais raster e vetoriais. <https://anaconda.org/conda-forge/gdal>
- [7] GTiff – GeoTIFF File Format <https://gdal.org/drivers/raster/gtiff.html>
- [8] Scikit-learn: Machine Learning in Python <https://scikit-learn.org/stable/>