

Inteligência Artificial – Análise de dados do desmatamento na Amanozia

Guilherme Araújo Mendes de Souza
Instituto de Ciência e Tecnologia
Universidade Federal de São Paulo
São José dos Campos – São Paulo
guilherme.araujo13@unifesp.br

Resumo - Este relatório apresenta quatro algoritmos de inteligência artificial dedicados à análise de dados obtidos através do Instituto Nacional de Pesquisas Espaciais - INPE sobre o desmatamento na Amazônia. A abordagem proposta envolve o treinamento dos quatro algoritmos em um mesmo conjunto de dados, permitindo assim a comparação de seus desempenhos. Além disso, é destacado a urgência em combater o desmatamento na região, ressaltando a capacidade desses algoritmos em visualizar padrões complexos. Foi evidenciando uma certa eficácia dos algoritmos nas suas aplicações. Adicionalmente, são discutidas estratégias para combater o desmatamento na Amazônia.

Abstract — This report presents four artificial intelligence algorithms dedicated to the analysis of data obtained through the National Institute for Space Research - INPE regarding deforestation in the Amazon rainforest. The proposed approach involves training the four algorithms on the same dataset, allowing for the comparison of their performances. Additionally, the urgency of combating deforestation in the region is emphasized, highlighting the ability of these algorithms to visualize complex patterns. The report demonstrates a certain effectiveness of the algorithms in their applications. Furthermore, strategies to combat deforestation in the Amazon are discussed.

I. INTRODUÇÃO

Na virada do século XX para o XXI, observamos um aumento alarmante no desmatamento na Amazônia, um fenômeno que transcende fronteiras e demanda uma atenção urgente das autoridades. As atividades predominantes, especialmente na Amazônia Brasileira, responsáveis por essa devastação, têm sua raiz na agropecuária, mineração ilegal e o desflorestamento ilegal alimentando um ciclo de degradação ambiental. Processos como incêndios florestais, mortalidade de árvores devido à seca e calor extremos, além da liberação de estoques de carbono no solo. Este cenário não apenas compromete a rica biodiversidade da região, mas também desencadeia consequências significativas para o equilíbrio ambiental global.

À medida que o desmatamento avança, suas repercussões tornam-se cada vez mais evidentes. A perda de serviços ambientais, incluindo a interrupção da ciclagem de água e a liberação de estoques de carbono essenciais para amenizar o agravamento do efeito estufa, surgem como uma preocupação crítica. Além de ferir diretamente o objetivo 13 – Ação contra a mudança global do clima dos 17 Objetivos de Desenvolvimento Sustentável para 2030, e indiretamente os objetivos 14 – Vida na água e 15 – Vida terrestre. Uma vez que o desmatamento possui impactos significativos nos ecossistemas aquáticos e que se não for gerido de forma

sustentável, pode levar à perda de biodiversidade e degradação dos ecossistemas terrestres.

As pesquisas sobre o desmatamento têm ganhado destaque devido à urgência do tema. Diversos estudos abordam diferentes aspectos, Estudos do Leandro Valle et al., a respeito do desmatamento na Amazônia e a importância das áreas protegidas. Enquanto Philip M. Fearnside do INPA sobre a dinâmica, impactos e controle do desmatamento na Amazônia. Sérgio Rivero, Oriana Almeida et al., da Universidade Federal do Pará contribuíram com a análise das causas diretas do desmatamento na Amazônia. Luana Helena, Paula Fernanda et al., que usaram a inteligência artificial para prognóstico de desmatamento na Rebio do Gurupi – MA. Além do Bruno J. T. Fernandes et al., da Universidade de Pernambuco avaliam técnicas de IA para previsão de incidência de desmatamento na Amazônia.

Este relatório visa explorar não apenas as causas e consequências do desmatamento na Amazônia, mas também apresentar abordagens, com algoritmos de inteligência artificial, na análise de dados para combater essa crise ambiental.

O combate ao desmatamento deve ser tratado como uma pauta de urgência pelo atual Governo. Nesse sentido, o governo e organizações não governamentais (ONGs) ambientais, tem que se uniram para adotar medidas contra o desmatamento. Como o aumento da fiscalização, através da análise de dados disponíveis no INPE. Dessa forma, ao intensificar a fiscalização e monitoramento por meio de dados e algoritmos de inteligência artificial, será possível detectar e combater o desmatamento. Possibilitando assim a preservação da Amazônia a longo prazo.

II. TRABALHOS RELACIONADOS

1. Leandro Valle Ferreira, Eduardo Venticinque e Samuel Almeida, “O desmatamento na Amazônia e a importância das áreas protegidas”, DOSSIÊ AMAZÔNIA BRASILEIRA I, 2005.

Esse artigo analisou os fatores que levaram à degradação ambiental da região e o papel das unidades de conservação e das terras indígenas na preservação da floresta. O artigo usa dados do INPE para medir o desmatamento nos estados de Mato Grosso, Rondônia e Pará, e mostrou que as áreas protegidas apresentam taxas de desmatamento muito menores do que as áreas não protegidas.

2. Philip M. Fearnside, “Desmatamento na Amazônia: dinâmica, impactos e controle”, Acta Amazonica - INPA, 2006.

O artigo levantou as causas e as consequências do desmatamento na região e as possíveis soluções para reduzi-lo. Foi utilizado dados do INPE para medir o desmatamento nos estados de Mato Grosso, Rondônia e Pará, e mostra que o desmatamento leva à perda de serviços ambientais. O artigo também discute os efeitos das mudanças climáticas e dos fenômenos de El Niño na floresta, e alerta para os riscos de retroalimentações entre o clima e a floresta, que podem levar a incêndios florestais, mortalidade de árvores e liberação de carbono no solo.

3. Sérgio Rivero, Oriana Almeida, Saulo Ávila e Wesley Oliveira, “Pecuária e desmatamento: uma análise das principais causas diretas do desmatamento na Amazônia”, Nova Economia - UFMG, 2009.

Esse artigo analisou as principais causas diretas do desmatamento na região, usando dados em painel. O artigo usa dados do INPE-PRODES para medir o desmatamento nos estados de Mato Grosso, Rondônia e Pará, e dados do IBGE para medir o número de cabeças de gado e a área plantada de culturas perenes e temporárias. O artigo mostra que o desmatamento é fortemente correlacionado com a pecuária, e que a soja também tem uma influência positiva no desmatamento.

4. Luana Helena de Oliveira Gama, Paulo Fernando Paiva Rodrigues Paiva, Osmar Martins da Silva-Júnior e José Salatiel Nahum, “Modelagem ambiental e uso da inteligência artificial para prognóstico de desmatamento: o caso da Rebio do Gurupi - MA”, Revista Research, Society and Development, 2021.

Esse artigo teve como objetivo modelar cenários futuros na Reserva Biológica do Gurupi-MA usando o software Dinamica EGO, usando o método de transição para simular as trajetórias de desmatamento até 2030, com base nas variáveis: altitude, declividade, estradas, assentamento e área hidrográfica. O trabalho usa dados do INPE-PRODES para medir o desmatamento na área de estudo, e dados do IBGE para obter as variáveis independentes. O artigo mostra que o desmatamento na Rebio do Gurupi-MA tem uma tendência de crescimento, e que as variáveis que mais influenciam o desmatamento são a distância das estradas e a declividade.

5. Bruno J. T. Fernandes, José Roberto Osses e Rafael Façanha, “Avaliação de técnicas de IA para auxiliar na previsão de incidência de desmatamento na Amazônia.”, Revista de Engenharia e Pesquisa Aplicada, 2022.

O artigo comparou o desempenho de diferentes algoritmos de classificação na análise de imagens de satélite da série histórica do INPE-PRODES, usando como variáveis de entrada a intensidade de desmatamento anterior, a distância de cursos de água e a distância de rodovias. O artigo usa dados

do INPE-PRODES para medir o desmatamento nos estados de Mato Grosso, Rondônia e Pará, e dados do IBGE para obter as variáveis independentes. O artigo mostra que os algoritmos de classificação conseguem prever com boa acurácia a incidência de desmatamento nas áreas de estudo, e que os algoritmos que apresentaram os melhores resultados foram o Random Forest e o Support Vector Machine.

III. METODOLOGIA

Este relatório tem como objetivo usar aprendizado de máquina para a classificação de padrões em um conjunto de dados referente ao desflorestamento na Amazonia entre os anos de 2000 e 2020.

A implementação foi realizada usando a biblioteca de aprendizado de máquina sklearn, na linguagem de programação Python. Os dados foram coletados no site do kaggle, retirados do Instituto Nacional de Pesquisas Espaciais (INPE). O conjunto de dados contém dados históricos de satélite.

Foi utilizado um conjunto de treinamento e um conjunto de teste para avaliar o desempenho do modelo, separando 70% dos dados para treinamento e 30% para validação.

O conjunto de treinamento foi utilizado para treinar os modelos de aprendizado de máquina usando os algoritmos KNN, SVM, mlp e Keras. Os dados foram pré-processados usando técnicas como normalização. O modelo foi então testado no conjunto de teste.

A variável dependente foi a class atribuída a cada elemento do dataframe, usando como parâmetro Área do município (area) / Área de desmatamento no ano (incremento).

Tabela 1 – Atribuição das class

class	Área desmatada (%)
1	100% - 10%
2	10% - 5%
3	5% - 3%
4	3% - 1%
5	1% - 0.7%
6	0.7% - 0.5%
7	0.5% - 0.3%
8	0.3% - 0.2%
9	0.2% - 0.1%
10	0.1% - 0.01%
11	Menores que 0.01%

IV. ANÁLISE EXPERIMENTAL

1. Conjunto de dados

1. Conjunto de dados

O dataset é formatado por três arquivos CSV (Comma-Separated values), o Counties que contém o Nome Microrregião e o Código Município Completo, states que contém os estados_id, Estados e o data que contém os dados principais como:

- Ano: O ano do registro;
- id_municipio: O código de identificação;
- Area: A área total do município;
- Desmatado: O histórico de desmatamento acumulado;
- Incremento: O desmatamento registrado no ano;
- Floresta: A área total composta por floresta no município
- Nuvem: A área da região do município, das imagens do satélite que foram obstruídas por nuvens;
- Nao_observado: A área que não possível observar;
- Nao_floresta: A área do município que não é composta por floresta;
- Hidrografia: A área do município que é composta por Rios, lagos ou represas.

2. Ambiente do algoritmo e ambiente computacional

Ambiente (hardware e software):

- CPU: Intel I5 10400;
 - o Clock: 2.90 GHz - 4.30 GHz;
 - o Cache: 12 MB;
 - o Núcleos: 6;
 - o Threads: 12;
- Memória: 16 GB 2133 MHz;
- GPU: NVIDIA GeForce GTX 1650 – 4 GB;
- Sistema Operacional: Windows 10 Pro – 22H2 - x64;
- Compilador: Visual Studio Code - Versão: 1.84.2;
- Linguagem: Python 3.11.6;

3. Critérios de análise

A avaliação do desempenho dos modelos foi realizada usando as métricas disponíveis na biblioteca sklearn, calculando a acurácia de cada modelo a partir da validação dos modelos nos conjuntos de teste e treinamento. Além disso foi utilizado o método `accuracy_score_with_tolerance`, uma função personalizada que calcula a acurácia dos modelos considerando uma tolerância de uma classe para cima e uma para baixo. Ou seja, se o modelo prevê uma classe 3 e o valor real é uma classe 2 ou 4, o modelo não é penalizado por esse erro. Essa abordagem é mais adequada para o problema em questão, pois a variação entre as porcentagens de cada classe é baixa e seria muito complexo o algoritmo acertar a classe exata de cada elemento.

4. Resultados

a) Leitura dos dados

Após a leitura e união dos datasets, foi realizado uma análise inicial do dataframe resultante. Este conjunto de dados contém 16720 linhas de dados e 12

colunas de features. Além verificar os tipos de dados, sendo eles int, float e object.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16720 entries, 0 to 16719
Data columns (total 12 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   ano          16720 non-null  int64
1   id_municipio 16720 non-null  int64
2   area         16720 non-null  int64
3   desmatado    16720 non-null  float64
4   incremento   15960 non-null  float64
5   floresta     16720 non-null  float64
6   nuvem        16720 non-null  float64
7   nao_observado 16720 non-null  float64
8   nao_floresta 16720 non-null  float64
9   hidrografia  16720 non-null  float64
10  municipios   16720 non-null  object
11  estados      16720 non-null  object
dtypes: float64(7), int64(3), object(2)
memory usage: 1.5+ MB
```

Fig. 1 - Análise inicial

Essa diversidade de tipos reflete a natureza divergente dos dados coletados. Ademais o número de colunas e a variedade de tipos de dados estabelecem uma base para as etapas seguintes de pré-processamento e análise exploratória.

Ao realizar uma análise no conjunto de dados consolidado, foi identificado a presença de valores ausentes. Esses dados inexistente foram observados na coluna de incremento, referentes ao ano de 2000.

```
ano          0
id_municipio 0
area         0
desmatado    0
incremento    760
floresta     0
nuvem        0
nao_observado 0
nao_floresta 0
hidrografia  0
municipios   0
estados      0
dtype: int64
```

Fig. 2 - Análise de valores ausentes

Considerando o contexto específico do conjunto de dados foi adotado a estratégia de substituir os valores ausentes por 0. Essa abordagem é justificada pelo fato de que era o primeiro ano de contagem dos dados.

```
ano          0
id_municipio 0
area         0
desmatado    0
incremento    0
floresta     0
nuvem        0
nao_observado 0
nao_floresta 0
hidrografia  0
municipios   0
estados      0
dtype: int64
```

Fig. 3 - Dataframe corrigido

b) Distribuição dos dados

Durante a análise da distribuição dos dados, observamos que os dados não apresentam uma distribuição não uniforme ou equilibrada. Essa observação é evidenciada ao comparar os elementos da classe 1 e da classe 11 (Fig.4).

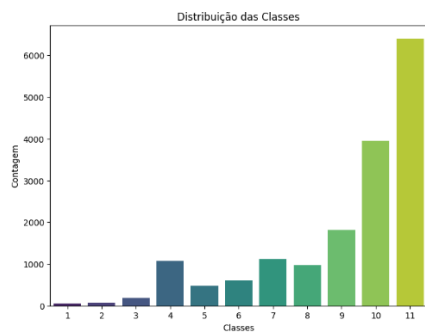


Fig. 4 - Distribuição das classes

c) Relação entre os dados

Através da matriz de correlação (Fig. 5) foi possível observar uma relação esperada entre area e floresta, também foi possível observar uma leve relação entre floresta e hidrografia, area e hidrografia, area e nuvem e desmatado e incremento.

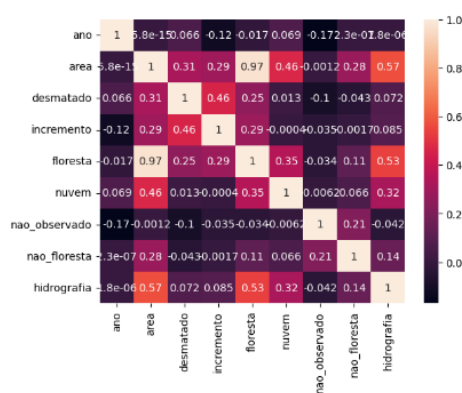


Fig. 5 - Matriz de correlação

d) Análise dos dados

A Figura 6 fornece uma visão temporal das taxas de desmatamento e florestamento ao longo dos anos. Uma observação notável é que, no ano de 2005, a taxa de desmatamento superou a taxa de florestamento.

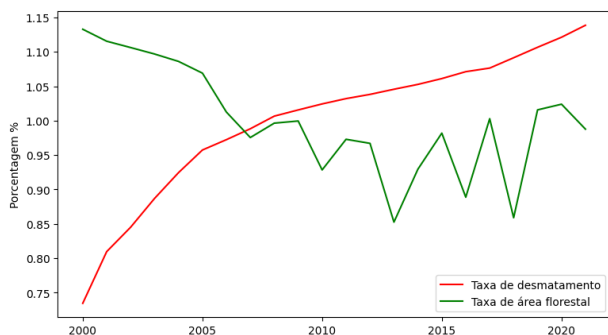


Fig. 6 - Comparação entre as taxas de desmatamento e a taxa de área florestal

O gráfico também evidencia um ponto crítico em que as áreas desmatadas superaram significativamente as áreas reflorestadas ao longo dos anos. Essa disparidade entre desmatamento e florestamento nesse período específico destaca um período intenso no desmatamento da Amazonia.

A figura 7 mostra a área aproximada de cada estado. Já a figura 8 mostra a área florestal de cada estado.

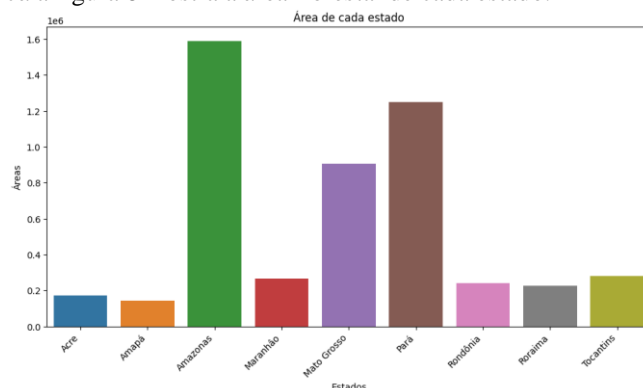


Fig. 7 – Área dos estados

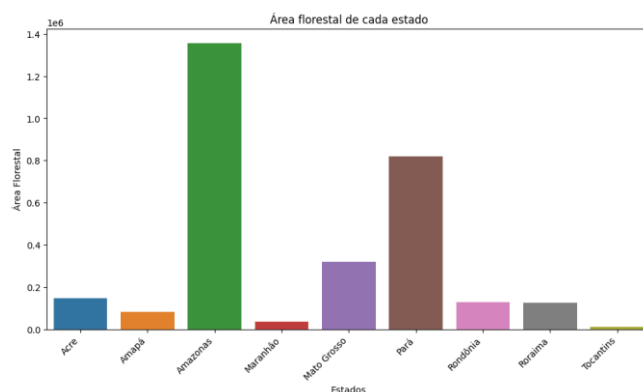


Fig. 8 – Área Florestal dos estados

Após análise dos dois gráficos, é possível os estados do Amazonas, Pará e Mato Grosso possuem maior área do Bioma. Também é possível verificar que os estados do Tocantins, Maranhão e Mato Grosso possuem um baixo nível florestal comparado a suas áreas.

A Figura 9 oferece uma representação visual do desmatamento ao longo dos anos, proporcionando uma perspectiva dinâmica das mudanças nas áreas desmatadas.

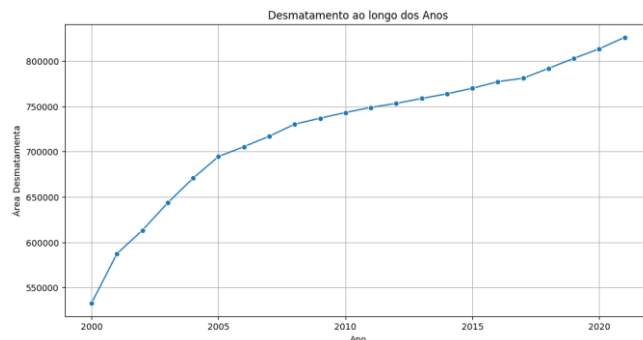


Fig. 9 – Área de desmatamento ao longo dos anos

A Figura 10 oferece uma representação visual do desmatamento ao longo dos anos, com um destaque particular para a crescente atividade de desmatamento nos estados do Pará e Mato Grosso.

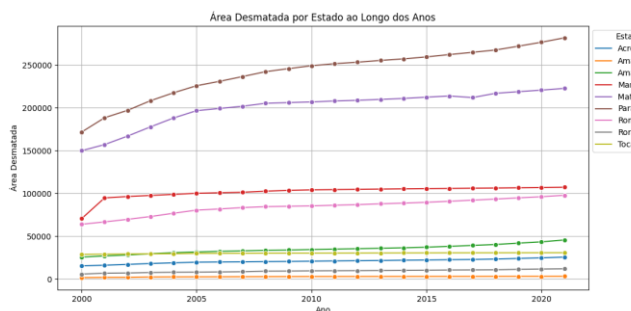


Fig. 10 – Área de desmatamento os estados ao longo dos anos

A Figura 11 apresenta uma visualização detalhada do desmatamento ao longo dos anos, com ênfase na crescente atividade nas cidades de São Felix do Xingu, Altamira e Porto Velho.

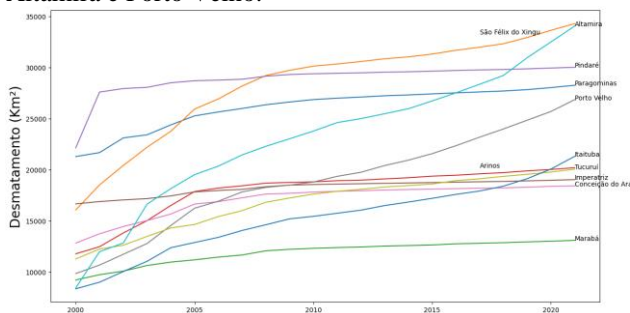


Fig. 11 – Área em Km² das cidades que mais desmataram

A análise revela uma tendência preocupante de aumento significativo no desmatamento nessas cidades ao longo dos anos.

e) Resultados dos algoritmos de aprendizado de máquina

Por fim foi realizado a avaliação do desempenho dos modelos aplicados, utilizando métricas de acurácia simples e acurácia com tolerância. Abaixo estão os resultados obtidos para cada modelo:

SVM:

Acurácia Simples: 50.28%

Acurácia com Tolerância: 72.34%

MLP:

Acurácia Simples: 48.78%

Acurácia com Tolerância: 71.47%

KNN:

Acurácia Simples: 53.75%

Acurácia com Tolerância: 78.84%

Keras:

Acurácia Simples: 53.40%

Acurácia com Tolerância: Não foi possível calcular.

Esses resultados fornecem clareza sobre a eficácia relativa dos modelos, destacando a importância de considerar a tolerância na avaliação do desempenho.

V. CONCLUSÃO

Neste relatório, foi implementado uma abordagem de aprendizado de máquina supervisionado para a classificação de padrões em um conjunto de dados proveniente do site do INPE. Os modelos, treinados utilizando os algoritmos KNN, SVM, MLP e Keras, foram avaliados quanto ao seu desempenho usando métricas disponíveis na biblioteca sklearn.

Os resultados indicam um desempenho razoável dos algoritmos, demonstrando uma certa efetividade na classificação de áreas desmatadas. A incorporação da função `accuracy_score_with_tolerance` proporcionou uma avaliação mais robusta, considerando tolerâncias para cima e para baixo na classificação.

As relações identificadas durante a análise dos dados e a execução dos modelos oferecem contribuições significativas no auxílio ao combate ao desmatamento. Destaca-se a relevância do aprendizado de máquina na análise de dados ambientais, fornecendo ferramentas eficazes para a identificação e classificação de áreas impactadas.

As relações encontradas durante a análise dos dados e durante a execução dos modelos contribuem de forma significativa no auxílio ao combate ao desmatamento destacando a relevância do aprendizado de máquina na análise de dados ambientais.

Apesar dos resultados promissores, reconheço que uma análise mais aprofundada seria possível com um dataset mais ricos em informações e dados mais específicos relacionados às causas do desmatamento. A inclusão de informações mais detalhadas poderia enriquecer a análise e permitir análises mais profundas.

Acredito que os modelos implementados, embora mostrem um desempenho razoável, sugerem a necessidade de um melhor tratamento nos dados ou até um de conjunto de dados para uma análise mais abrangente e precisa.

VI. BIBLIOGRAFIA

L. V. Ferreira, E. Venticinque e S. Almeida, “O desmatamento na Amazônia e a importância das áreas protegidas”, *Estud. Avancados*, vol. 19, n.º 53, pp. 157–166, abril de 2005. Consult. 2023-12-01. [Em linha]. Disponível: <https://doi.org/10.1590/s0103-40142005000100010>

P. M. Fearnside, “Desmatamento na Amazônia: dinâmica, impactos e controle”, *Acta Amaz.*, vol. 36, n.º 3, pp. 395–400, 2006. Consult. 2023-12-01. [Em linha]. Disponível: <https://doi.org/10.1590/s0044-59672006000300018>

S. Rivero, O. Almeida, S. Ávila e W. Oliveira, “Pecuária e desmatamento: uma análise das principais causas diretas do desmatamento na Amazônia”, *Nova Econ.*, vol. 19, n.º 1, pp. 41–66, 2009. Consult. 2023-12-01. [Em linha]. Disponível: <https://doi.org/10.1590/s0103-63512009000100003>

B. J. T. Fernandes, J. R. Osses e R. Façanha, “Avaliação de Técnicas de IA para Auxiliar na Previsão de Incidência de Desmatamento na Amazônia”, Rev. Eng. Pesqui. Apl., vol. 7, n.º 2, pp. 57–64, julho de 2022. Consult. 2023-12-01. [Em linha]. Disponível: <https://doi.org/10.25286/rep.v7i2.2218>

L. H. O. M. Gama, P. F. P. R. Paiva, O. M. d. Silva Junior e M. d. L. P. Ruivo, “Modelagem ambiental e uso da inteligência artificial para prognóstico de desmatamento: o caso da Rebio do Gurupi-MA”, Res., Soc. Develop., vol. 10, n.º 2, fevereiro de 2021, art. n.º e13810211609. Consult. 2023-12-01. [Em linha]. Disponível: <https://doi.org/10.33448/rsd-v10i2.11609>

“Amazônia Rainforest Deforestation 2000 - 2021”. Kaggle: Your Machine Learning and Data Science Community. Consult. 2023-11-28. [Em linha]. Disponível: <https://www.kaggle.com/datasets/diegosilvadefrana/brazilian-deforestation-from-2000-to-2021>

alvaropagliari. “Amazônia Deforestation Forecast”. Kaggle: Your Machine Learning and Data Science Community. Consult. 2023-11-28. [Em linha]. Disponível: <https://www.kaggle.com/code/alvaropagliari/amazonia-deforestation-forecast>

L. R. Catunda. “Analisando dados de desmatamento da Amazônia com Python e Pandas”. Medium. Consult. 2023-12-01. [Em linha]. Disponível: <https://medium.com/@Regisc/analizando-dados-de-desmatamento-da-amazonia-com-python-e-pandas-afbf713c2ddb>