

**UNIVERSIDADE DE SÃO PAULO**

Instituto de Ciências Matemáticas e de Computação

## **Análise do desmatamento na Amazônia brasileira: uma aplicação de aprendizado de máquina**

**Fernanda Coff Dias**

Monografia - MBA em Inteligência Artificial e Big Data



SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: \_\_\_\_\_

**Fernanda Coff Dias**

## **Análise do desmatamento na Amazônia brasileira: uma aplicação de aprendizado de máquina**

Monografia apresentada ao Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, como parte dos requisitos para obtenção do título de Especialista em Inteligência Artificial e Big Data.

Área de concentração: Inteligência Artificial

Orientadora: Profa. Dra. Heloisa de Arruda Camargo

**Versão original**

**São Carlos**

**2023**

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi  
e Seção Técnica de Informática, ICMC/USP,  
com os dados inseridos pelo(a) autor(a)

D674a      Dias, Fernanda Coff  
            Análise do desmatamento na Amazônia brasileira:  
            uma aplicação de aprendizado de máquina / Fernanda  
            Coff Dias; orientadora Heloisa de Arruda Camargo. -  
            - São Carlos, 2023.  
            58 p.

            Trabalho de conclusão de curso (MBA em  
            Inteligência Artificial e Big Data) -- Instituto de  
            Ciências Matemáticas e de Computação, Universidade  
            de São Paulo, 2023.

            1. Desmatamento. 2. Amazônia. 3. Aprendizado de  
            Máquina. I. Camargo, Heloisa de Arruda , orient.  
            II. Título.

**Fernanda Coff Dias**

Monograph presented to the Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, as part of the requirements for obtaining the title of Specialist in Artificial Intelligence and Big Data.

Concentration area: Artificial Intelligence

**Original version**

**São Carlos  
2023**

## RESUMO

DIAS, F. C. **Análise do desmatamento na Amazônia brasileira: uma aplicação de aprendizado de máquina.** 2023. 58p. Monografia (MBA em Inteligência Artificial e Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2023.

As principais causas do desmatamento na Amazônia vêm sendo estudadas a partir das bases de dados cada vez mais extensas disponíveis. Entre as principais causas do desmatamento destacadas na literatura estão a agricultura e a pecuária. Utilizam-se, principalmente, para fazer essas análises, modelos lineares de regressão e, mais recentemente, algoritmos de aprendizado de máquina, em especial modelos baseados em árvore. O objetivo deste trabalho é obter um modelo preditivo para a taxa anual de desmatamento na Amazônia em  $\text{km}^2$ , a partir de quinze variáveis explicativas socioeconômicas utilizando-se modelos de aprendizado de máquina. Quatro modelos de árvore, a saber, Árvore de Decisão, Floresta Aleatória, Gradient Boosting e Extra Trees, e o modelo Máquinas de Vetor-Suporte foram avaliados. Dados de desmatamento do INPE-PRODES e dados do IBGE, Instituto Sócio Ambiental e Wikipedia foram utilizados na análise. Entre as variáveis utilizadas destacaram-se como mais importantes para o processo preditivo dos modelos as áreas colhidas de lavouras permanente e temporárias. O modelo com melhor desempenho em termos das métricas coeficiente de determinação e erro quadrático médio foi o Floresta Aleatória. O algoritmo Gradient Boosting foi o que resultou nas piores valores previstos para a variável alvo.

**Palavras-chave:** Desmatamento, Amazônia, aprendizado de máquina.

## **ABSTRACT**

DIAS, F. C. **Analysis of deforestation in the Brazilian Amazon: an application of machine learning.**. 2023. 58p. Monograph (MBA in Artificial Intelligence and Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2023.

The main causes of deforestation in the Brazilian Amazon have been studied based on the increasingly extensive databases available. Among the main causes highlighted in the literature are agriculture and Cattle Ranching. To carry out these analyses, linear regression models and, more recently, machine learning algorithms, in special tree-based models, are traditionally used. The objective of this work is to obtain a predictive model for deforestation in the Brazilian Amazon, based on fifteen socioeconomic explanatory variables using machine learning regression models. Four tree based models were evaluated: Decision Tree, Random Forest, Gradient Boosting and Extra Trees, and the Support Vector Machines algorithm. Deforestation data from INPE-PRODES, and data from IBGE, Instituto Sócio Ambiental and Wikipedia were used in the analysis. Among the variables used, the areas of temporary and permanent crops were the most important for the predictive power of the models. The model with the best performance in terms of the coefficient of determination and mean squared error metrics was the Random Forest. The Gradient Boosting algorithm resulted in the worst predicted values for the target variable.

**Keywords:** Deforestation, Amazon, machine learning.





## LISTA DE ABREVIATURAS E SIGLAS

AM	Aprendizado de Máquina
AD	Árvore de Decisão (algoritmo)
BCB	Banco Central do Brasil
DETER	Detecção de Desmatamento em Tempo Real
ET	Extra Tree (algoritmo)
FA	Floresta Aleatória (algoritmo)
GB	Gradient Boosting (algoritmo)
IBGE	Instituto Brasileiro de Geografia e Estatística
INPE	Instituto Brasileiro de Pesquisas Espaciais
IPEA	Instituto de Pesquisa Econômica Aplicada
ISA	Instituto SocioAmbiental
MSE	Erro quadrático médio
MAE	Erro absoluto médio
PIB	Produto Interno Bruto
PRODES	Programa de monitoramento contínuo do desmatamento
SVM	Máquinas de Vetores-Suporte (algoritmo)
UCs	Unidades de Conservação



## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>13</b>
<b>2</b>	<b>REVISÃO DE LITERATURA</b>	<b>15</b>
<b>2.1</b>	<b>Sensoriamento remoto por satélites</b>	<b>15</b>
<b>2.2</b>	<b>Métodos estatísticos e de aprendizado de máquina para estudo dos dados disponibilizados pelo INPE</b>	<b>15</b>
<b>3</b>	<b>PRÉ-PROCESSAMENTO E ANÁLISE EXPLORATÓRIA DOS DADOS</b>	<b>19</b>
<b>3.1</b>	<b>Área de estudo</b>	<b>19</b>
<b>3.2</b>	<b>Visão Geral</b>	<b>19</b>
<b>3.3</b>	<b>Pré-processamento dos dados</b>	<b>20</b>
<b>3.4</b>	<b>Análise exploratória dos dados</b>	<b>21</b>
3.4.1	Desmatamento	21
3.4.2	Queimadas	22
3.4.3	PIB	25
3.4.4	PIB Estadual Agropecuário	25
3.4.5	Área plantada ou destinada à colheita	26
3.4.6	Valor da produção agropecuária	27
3.4.7	População	28
3.4.8	Fluxo de crédito rural	28
3.4.9	Extensão de UCs federais e estaduais	29
3.4.10	Efetivo Bovino	31
3.4.11	Área colhida de lavoura permanente e temporária	31
3.4.12	Produção de Milho e Soja	33
3.4.13	Despesa orçamentária por função - gestão ambiental, agricultura e organização agrária	34
3.4.14	Partido do Governador em Exercício	35
3.4.15	Matriz de correlação	35
<b>4</b>	<b>MÉTODOS</b>	<b>41</b>
<b>4.1</b>	<b>Algoritmos</b>	<b>41</b>
4.1.1	Árvores de Decisão	41
4.1.1.1	Floresta Aleatória e Extra Trees	41
4.1.2	Gradient Boosting	42
4.1.3	Máquinas de vetores-suporte	42

<b>4.2</b>	<b>Medidas de desempenho</b>	<b>42</b>
4.2.1	Coeficiente de determinação ( $R^2$ )	42
4.2.2	Erro quadrático médio (MSE)	43
4.2.3	Erro absoluto médio (MAE)	43
<b>5</b>	<b>RESULTADOS DOS EXPERIMENTOS</b>	<b>45</b>
<b>5.1</b>	<b>Desenho dos Experimentos</b>	<b>45</b>
<b>5.2</b>	<b>Resultados Gerais</b>	<b>45</b>
5.2.1	Modelos de aprendizado de máquina	48
<b>6</b>	<b>CONCLUSÃO</b>	<b>53</b>
	<b>REFERÊNCIAS</b>	<b>55</b>

## 1 INTRODUÇÃO

Muitos são os fatores atribuídos ao crescimento do desmatamento na Amazônia brasileira. Dentre as causas consideradas mais relevantes estão a pecuária, a agricultura, a extração de madeira (RIVERO *et al.*, 2009), o aumento populacional e as políticas governamentais como incentivo à exportação e crédito rural (YOUNG, 2022).

Com o objetivo de entender as causas do desmatamento, promover ações de controle e fiscalização, compreender processos ambientais e transformações climáticas, prever ocorrências de incêndios florestais, entre outras aplicações, a existência de dados consistentes e atualizados é fundamental. Desde a década de 1970, informações precisas e constantes obtidas a partir de sensoriamento remoto por satélites vêm sendo coletadas em todo o planeta (CHUVIECO *et al.*, 2019; WOOSTER *et al.*, 2021; HITCHCOCK; HOFFER, 1974). O sensoriamento remoto é o único meio viável e eficaz de monitorar contínua e regularmente o estado e a dinâmica de grandes áreas de florestas (CHEN *et al.*, 2016; COHEN; GOWARD, 2004). Os dados cada vez mais precisos e em maior número obtidos por satélites, combinados com a adequada manipulação e o avanço nas técnicas de aprendizagem de máquina são elementos essenciais para o desenvolvimento e aperfeiçoamento dos sistemas de monitoramento do clima, queimadas e desmatamento.

No Brasil, o Instituto Nacional de Pesquisas Espaciais (INPE) realiza o monitoramento do desmatamento e dos incêndios florestais na Amazônia através dos projetos PRODES, QUEIMADAS e DETER. Através do PRODES, o INPE divulga anualmente as taxas de desmatamento desde 1988. São utilizadas neste programa imagens dos sensores dos satélite LANDSAT da NASA, satélites da Disaster Monitoring Constellation e satélites CBERS do INPE, que cobrem a Amazônia com baixa frequência temporal (16 e 26 dias) e têm resolução espacial de 30 e 20 metros, respectivamente, possibilitando o mapeamento de desmatamentos cujas áreas sejam superiores a 6,25 hectares. O PRODES identifica e contabiliza apenas áreas de corte raso da floresta, que são aquelas que estão no estágio final do processo de desmatamento (INPE, b).

Uma vez que os dados de desmatamento do PRODES são divulgados uma vez por ano e demoram para ser processados, o DETER-A e posteriormente DETER-B (Detecção de Desmatamento em Tempo Real) foi criado para que fossem dadas respostas mais rápidas para ações de fiscalização e controle. Neste sistema, utilizam-se dados de satélites que cobrem a Amazônia com maior frequência, mas cujas imagens têm menor resolução espacial (satélite CBERS-4 do INPE e IRS (Indian Remote SensingSatellite)). Além disso, monitora-se qualquer alteração na cobertura vegetal (INPE, b).

Além dos programas citados acima, desde 1985 o INPE desenvolve metodologias e

programas para identificar focos de calor nas imagens de satélites, através do programa QUEIMADAS. Os dados são obtidos a partir de satélites polares que fornecem pelo menos dois conjuntos de imagens por dia, e satélites geoestacionários que geram seis imagens por hora, sendo que no total o INPE processa automaticamente mais de 200 imagens por dia especificamente para detectar focos de queima da vegetação (INPE, c; INPE, a).

Os dados do desmatamento são utilizados na literatura para analisar a correlação entre desmatamento e diversas variáveis ambientais, climáticas e socioeconômicas. Tradicionalmente, utilizam-se métodos estatísticos, como regressão linear simples (JAKIMOW *et al.*, 2023; CÂMARA *et al.*, 2023; VOLPATO *et al.*, 2022; WANG; HUANG, 2022; RIVERO *et al.*, 2009). Técnicas de Aprendizado de Máquina (AM) também vêm sendo usadas para prever a probabilidade de ocorrência de incêndios florestais pelo uso de modelos como redes neurais artificiais e Floresta Aleatória (KIM *et al.*, 2019; OLIVEIRA *et al.*, 2012; VILAR *et al.*, 2016; LEUENBERGER *et al.*, 2018; THACH *et al.*, 2018). Esses algoritmos apresentam bom desempenho ao relacionar fenômenos não lineares e complexos, como é caso das variáveis envolvidas no princípio e propagação dos incêndios florestais.

O objetivo deste trabalho é analisar dados históricos e aplicar metodologias para gerar um modelo preditivo para o desmatamento na Amazônia brasileira. Para tanto, serão utilizados e comparados alguns algoritmos de aprendizado de máquina para regressão, avaliando-se o desempenho preditivo. Também será feita uma análise de correlação entre diversas variáveis, identificando-se assim aquelas que forem mais relevantes para o aumento das taxas de desmatamento. As análises serão realizadas a partir dos dados disponibilizados pelo INPE.

## 2 REVISÃO DE LITERATURA

Neste capítulo são apresentados os resultados da pesquisa de literatura sobre sensoriamento remoto por satélites e as principais publicações relativas ao estudo do desmatamento relacionado à diferentes variáveis socioeconômicas e ambientais abordadas a partir de diferentes métodos.

### 2.1 Sensoriamento remoto por satélites

Na técnica de sensoriamento remoto as informações sobre um objeto são obtidas à distância a partir de satélites equipados com sensores capazes de emitir e receber ondas eletromagnéticas (IBGE, 2023b). No estudo da superfície planetária são utilizados diversos sensores, cada um capaz de captar uma parte do espectro eletromagnético. As propriedades térmicas da superfície do planeta, por exemplo, são medidas por sensores térmicos na faixa do infravermelho (ELACHI; ZYL, 2021). Não é objetivo deste trabalho descrever o funcionamento dos sensores ou como são tratadas as imagens obtidas por eles.

O primeiro satélite capaz de monitorar recursos terrestres foi lançado pelos Estados Unidos através da agência espacial NASA em 1972, chamado de Landsat 1. Hoje encontram-se em operação os satélites Landsat 8 e 9 (NASA, 2023). Desde o lançamento destes satélites, houve um grande progresso no estudo dos recursos naturais e do clima terrestre.

Atualmente existem diversos satélites de sensoriamento remoto em operação. De acordo com o INPE, para o monitoramento de focos de incêndio, o Brasil processa imagens de dez satélites equipados com sensores térmicos, são eles NOAA-18, 19 e 20, METOP-B e C, NASA TERRA e AQUA, NPP-Suomi, GOES-16 e MSG-3, todos pertencentes a agências norte-americanas e europeias (INPE, c). Para o monitoramento do desmatamento são utilizados satélites Landsat da Nasa, CBERS do INPE e IRS da Índia. Duas antenas localizadas em território brasileiro são responsáveis por captar os sinais desses satélites, localizadas nos estados de Mato Grosso e São Paulo. Para os objetivos deste trabalho, não serão tratadas as imagens obtidas pelos satélites, apenas os dados resultantes do tratamento das imagens realizado previamente pelo INPE serão utilizados.

### 2.2 Métodos estatísticos e de aprendizado de máquina para estudo dos dados disponibilizados pelo INPE

Os trabalhos listados abaixo utilizam os dados do INPE (BDQUEIMADAS ou PRODES) para analisar e relacionar ocorrência de incêndios e taxa desmatamento com diferentes variáveis, através de diferentes métodos estatísticos e de aprendizado de máquina. Os trabalhos discutidos a seguir mostram o potencial de pesquisa dos dados disponibi-

lizados nessas plataformas para analisar, a partir de diferentes abordagens, as possíveis consequências e contribuições para o aumento de focos de incêndio e do desmatamento no país.

Jakimow *et al.* (2023) analisaram e compararam estatisticamente o histórico de desmatamento e incêndios entre os anos de 2014 a 2020 em uma área de 150000 km<sup>2</sup> ao sul do Pará e avaliaram os efeitos das políticas implementadas por cada governo no desmatamento e incêndios em termos de posse de terra e tipos de fazendas. Eles identificaram que o aumento na área queimada foi maior a partir do ano 2018, nos assentamentos agrários e unidades de conservação e nas propriedades rurais de médio e grande porte. As taxas de desmatamento foram maiores em terras públicas não designadas, seguidas das unidades de conservação. Os autores também observaram que maior parte das áreas queimadas naqueles anos foram em áreas já desmatadas.

Câmara *et al.* (2023) relacionaram os dados de desmatamento com a posse e o uso da terra entre os anos de 2008 e 2021 na Amazônia brasileira. Os autores identificaram que o desmatamento foi maior em áreas privadas e que maior parte é desmatamento ilegal. Aproximadamente 1% das propriedades privadas responderam a cerca de 80% das florestas derrubadas no período.

Volpato *et al.* (2022) utilizaram métodos estatísticos para analisar as ocorrências de focos de incêndio no estado de Mato Grosso entre os anos de 2001 a 2015 e relacionar com variáveis climáticas (precipitação, velocidade média do vento, temperatura média e umidade relativa do ar), parâmetros de vegetação e uso do solo. Em relação ao uso do solo, os autores identificaram que os focos de incêndio ocorreram em maior parte na estação seca, em áreas de mineração e áreas urbanas. Em relação à cobertura de vegetação, os focos se concentram em áreas de pastagem. Os autores encontraram uma correlação negativa entre focos de incêndio e precipitação e umidade relativa do ar.

Wang e Huang (2022) analisaram e relacionaram as emissões devido às queimadas e a poluição por aerossóis no sul da Amazônia e no Pantanal aos incêndios em áreas desmatadas e não desmatadas durante o período seco entre os anos de 2003 e 2019. A quantidade de carbono queimado na região não desmatada foi mais de dez vezes maior que na região desmatada e diminuiu significativamente durante a estação seca no período. Os autores encontraram uma diminuição na tendência de emissões devido à queimadas em áreas não desmatadas. A diminuição das emissões de incêndios está relacionada com as condições climáticas desfavoráveis para as emissões de incêndios nos anos analisados, incluindo o aumento da umidade e a velocidade mais lenta do vento à superfície. Estes resultados sugerem que o desmatamento não foi o principal fator na diminuição das emissões de incêndios durante o período 2003-2019.

Rivero *et al.* (2009) reuniram os dados de 782 municípios da Amazônia Legal brasileira para identificar as principais causas diretas do desmatamento. Os autores



utilizaram regressões lineares múltiplas sobre os dados do INPE, IBGE e Ipeadata entre os anos de 2000 e 2006. Os autores identificaram que a expansão da pecuária é a principal causa imediata do processo de desmatamento e a soja impacta no desmatamento de forma indireta.

das Neves *et al.* (2021) utilizaram o algoritmo de aprendizado de máquina Floresta Aleatória para investigar a influência da rede rodoviária clandestina e regular no desmatamento da floresta amazônica e desenvolver uma abordagem para previsão do desmatamento utilizando uma série temporal de 1988 a 2018 no estado de Pará. As variáveis consideradas para relacionar com a infraestrutura rodoviária foram área desmatada (dados do INPE), população, renda familiar per-capita, taxa de desemprego, taxa de homicídios, PIB, exportação em valores monetários e percentual de analfabetos. Os autores encontraram que a rede rodoviária clandestina foi a variável com maior importância para o desmatamento, seguida da rede rodoviária regular. A curva de área desmatada ajustada pelo algoritmo Floresta Aleatória apresentou tendência consistente de crescimento do desmatamento até 2050, com taxas crescentes a cada 10 anos variando entre 4,64% e 11,54%, com média de 0,80% ao ano.



### 3 PRÉ-PROCESSAMENTO E ANÁLISE EXPLORATÓRIA DOS DADOS

Neste capítulo é apresentada uma visão geral sobre os dados selecionados para a análise. Em seguida, é descrito como foram tratados os dados. Por último, é feita a análise exploratória do conjunto de dados.

#### 3.1 Área de estudo

A região de estudo deste trabalho compreende o Bioma Amazônia Brasileiro que ocupa uma área de 4.196.943 km<sup>2</sup>, o que corresponde a mais de 40% do território nacional (IBF, 2020). Ele está presente nos estados do Acre, Amapá, Amazonas, Pará, Rondônia, Roraima e Tocantins, além de parte dos territórios do Mato Grosso e Maranhão. Para simplificar a construção do conjunto de dados, foram considerados os dados para todo o território destes últimos Estados, ainda que, possivelmente, não corresponda ao bioma Amazônia.

O clima predominante na região é o equatorial úmido, caracterizado pelo volume abundante de chuvas e temperaturas elevadas, variando entre 22 °C e 28 °C (IBF, 2020). A estação das chuvas ocorre entre os meses de novembro e abril.

#### 3.2 Visão Geral

Foram selecionados para analisar e relacionar aos dados de desmatamento na região Amazônica brasileira diferentes variáveis socioeconômicas cujos conjuntos de dados completos abrangessem o mesmo período dos dados do programa QUEIMADAS INPE (1999 a 2020) e que tivessem frequência anual. A Tabela 1 apresenta as variáveis escolhidas.

Todos os dados foram obtidos a partir de plataformas oficiais do governo federal (INPE e Instituto de Pesquisa Econômica Aplicada (IPEA), este último reúne dados obtidos ou reunidos por diferentes órgãos do governo como IBGE e Banco Central (BCB), na plataforma IPEADATA), do Instituto SocioAmbiental (ISA) e Wikipédia. Com exceção dos dados de focos de incêndio, os demais conjuntos de dados nos seus formatos originais abrangem períodos maiores, porém foram reduzidos ao período disponível para os dados de incêndios do INPE, começando em 1999. O ano de 2020 foi selecionado como limite superior das séries em função dos dados do PIB, População e Fluxo de crédito rural, que na plataforma IPEDATA estavam disponíveis somente até este ano.

Os dados finais abrangem, portanto, o período total de vinte e dois anos para os nove estados que possuem o bioma Amazônia nos seus territórios e 16 variáveis, formando um conjunto de 18 colunas e 198 linhas.

Tabela 1 – Variáveis utilizadas

Variáveis	Unidade	Tipo	Fonte
Desmatamento (alvo)	km <sup>2</sup>	Numérico (inteiro)	PRODES (INPE, b)
Focos de queimadas	Focos	Numérico (inteiro)	BDQUEIMADAS
PIB	R\$, a preços do ano 2010 (mil)	Numérico (real)	IBGE
Área plantada ou destinada à colheita	Hectare	Numérico (inteiro)	IBGE
Valor da produção agropecuária	R\$, a preços do ano 2010 (mil)	Numérico (real)	IBGE
PIB Estadual agropecuário	R\$, a preços do ano 2010 (mil)	Numérico (real)	IBGE
População	Pessoa	Numérico (inteiro)	IBGE
Fluxo de crédito rural	R\$, a preços do ano 2010	Numérico (real)	BCB
Extensão de UCs federais e estaduais	Hectares	Numérico (inteiro)	ISA
Efetivo Bovino	Cabeça	Numérico (inteiro)	IBGE
Área colhida de lavoura permanente	Hectare	Numérico (inteiro)	IBGE
Área colhida de lavoura temporária	Hectare	Numérico (inteiro)	IBGE
Produção de Milho	R\$, a preços do ano 2010 (mil)	Numérico (real)	IBGE
Produção de Soja	R\$, a preços do ano 2010 (mil)	Numérico (real)	IBGE
Despesa orçamentária por função - gestão ambiental, agricultura e organização agrária	R\$	Numérico (inteiro)	Ministério da Fazenda
Partido do Governador em exercício		String	Wikipedia

Fonte: Elaborada pela autora.

### 3.3 Pré-processamento dos dados

As bases de dados originais foram reunidas em uma única base de dados no formato de painel. Dados em painel são uma coleção de dados de vários objetos iguais ao longo do tempo, uma vez que tem-se várias variáveis para cada estado em cada ano. A Figura 1 apresenta uma visão das primeiras linhas e colunas dos dados no formato painel. Em destaque na Figura 1 estão as primeiras duas colunas de índices, o primeiro é o “Estado” e o segundo o “Ano”, uma vez que, para dados em painel, tem-se conjuntos de dados com mais de um índice (*MultiIndex*) (SHEPPARD, 2021).

Foram desconsiderados dos conjuntos de dados originais apenas os anos fora do

Figura 1 – Amostra dos dados em painel.

	Desmatamento	Extensao_de_Ucs	Fluxo_de_credito_rural	Populacao	PIB_Agropecuario	Valor_da_producao_agropecuaria	Area_plantada_ou_destinada	Focos_de_incendio	PIB	Efetivo_bovino	..
Estado	Ano										
AC	1999	441	2625745	4.634348e+06	559848	7.373892e+02	289.588476	91787	347	4.649466e+06	929999
	2000	547	2783339	7.146287e+07	576223	7.657779e+03	31364.569832	15167	43	4.845555e+04	1033311
	2001	419	3035220	6.137313e+06	5927	9.348420e+03	239926.837545	1264	829	5.244653e+05	1672598
	2002	883	3809418	6.773566e+07	586942	5.416837e+04	269826.593361	113139	6739	5.216824e+04	1817467
	2003	178	3809418	1.143757e+08	6595	6.349816e+04	3532.268146	113362	1523	5.165821e+06	1874804
	2004	728	4956185	4.627194e+07	63328	5.325942e+05	233521.378200	132684	7271	5.342621e+06	2062690
	2005	592	5282675	6.459799e+06	669736	5.917233e+04	254467.854119	134582	15993	5.625711e+06	2313185
	2006	398	5282675	3.523299e+07	686652	4.848615e+04	192953.685449	126861	6198	5.653945e+05	2452915
	2007	184	5282675	8.123879e+07	69999	4.814699e+05	114561.484315	123843	8549	6.269673e+06	2315798
	2008	254	5282675	6.664680e+04	6875	5.825133e+04	129468.926293	11355	5699	6.813460e+04	2425687
	2009	167	5308810	8.898584e+07	691132	5.449499e+05	94691.995564	94431	3511	7.356397e+04	2511285
	2010	259	5308810	1.335123e+08	734445	7.348640e+04	185599.619779	128842	8661	7.564554e+06	2578460
	2011	28	5308810	9.576359e+07	746386	7.189824e+04	372322.973588	13221	3191	7.538215e+06	2549497
	2012	35	5308810	3.647115e+07	758786	7.421888e+04	37196.575249	146928	472	7.825168e+06	2634467
	2013	221	5308810	1.968887e+08	776463	8.541449e+04	448795.125174	13776	498	8.343132e+05	2697489
	2014	39	5308810	2.897686e+07	7911	8.766939e+04	486965.643530	123511	4398	9.196385e+04	2799673
	2015	264	5308810	1.541336e+07	83513	8.357592e+05	376911.329854	116587	5779	8.533399e+06	2916207
	2016	372	5308810	1.455956e+08	816687	8.569484e+05	367735.336325	17294	7684	7.929818e+05	2998969
	2017	257	5464706	1.319917e+07	829619	7.598416e+05	314997.155942	1543	6295	7.854377e+06	2895028
	2018	444	5464706	1.783217e+07	869265	7.951588e+04	267881.586170	8533	6626	7.977137e+05	3303633
	2019	682	5464706	1.638448e+08	881935	5.871295e+05	23825.485777	8193	682	7.832572e+05	3509682
	2020	76	5464706	2.156871e+06	92442	5.192243e+04	232947.253518	8155	9193	7.882887e+05	3802112
AM	1999	72	16392636	3.658295e+07	2771785	4.533987e+03	623336.248277	213964	11	3.311787e+06	826025
	2000	612	16392636	1.143686e+08	2848785	3.923179e+04	585857.527968	29183	253	3.571115e+07	843254

Fonte: Elaborado pela autora.

período desejado. No período escolhido, os dados estavam completos e foram utilizados em seu formato original.

A variável “Partido político” foi transformada em variável binária com valores possíveis 0 ou 1, através do método *get\_dummies* da biblioteca Pandas. Após esta etapa, o conjunto de dados passa a ter 34 colunas e 198 linhas, ou seja, 16 colunas a mais.

Foi realizado o processo de normalização dos dados, de modo que cada atributo esteja no intervalo  $[0, 1]$ , através do método *MinMaxScaler*, disponível na biblioteca *scikit-learn* versão 1.2.0.

### 3.4 Análise exploratória dos dados

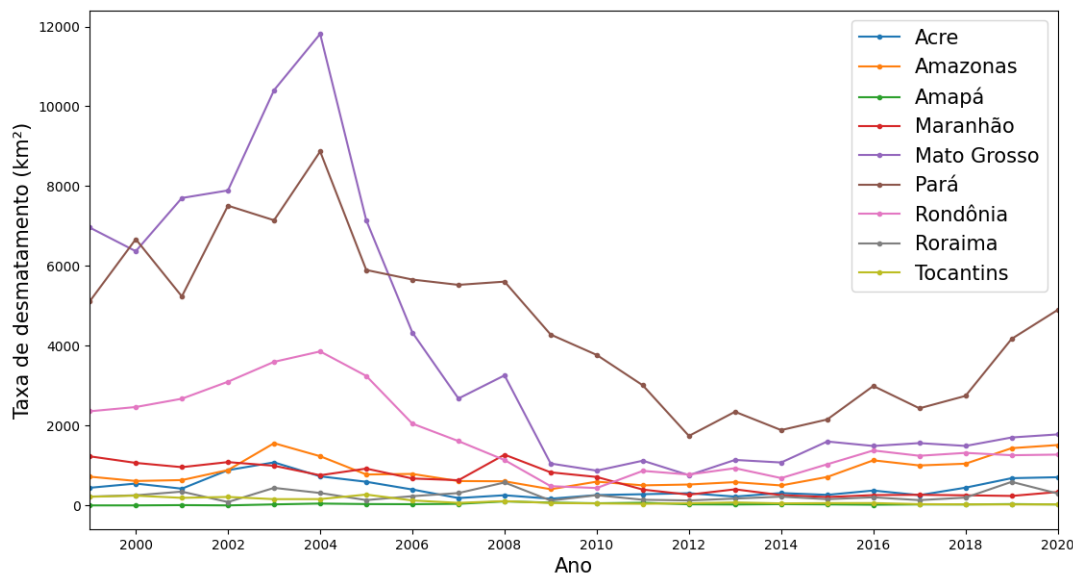
A seguir são detalhadas as informações sobre os conjuntos de dados apresentados na Tabela 1.

#### 3.4.1 Desmatamento

Os dados para o desmatamento foram obtidos através da página web do projeto PRODES (INPE, b). Analisando-se as taxas anuais de desmatamento por estado da região amazônica (INPE, b) nos anos de 1999 a 2020, Figura 2, observa-se que os estados do Pará e Mato Grosso são os que registraram maiores taxas de desmatamento para todos os anos desta série histórica. A taxa de área desmatada teve os maiores valores observados no ano de 2004. O estado do Pará registrou um aumento expressivo na área desmatada a

partir de 2018, quadruplicando a taxa de desmatamento entre este ano e o ano de 2020.

Figura 2 – Taxa anual de desmatamento por Estado do Bioma Amazônia.



Fonte: Elaborado pela autora com base nos dados do PRODES (INPE, b).

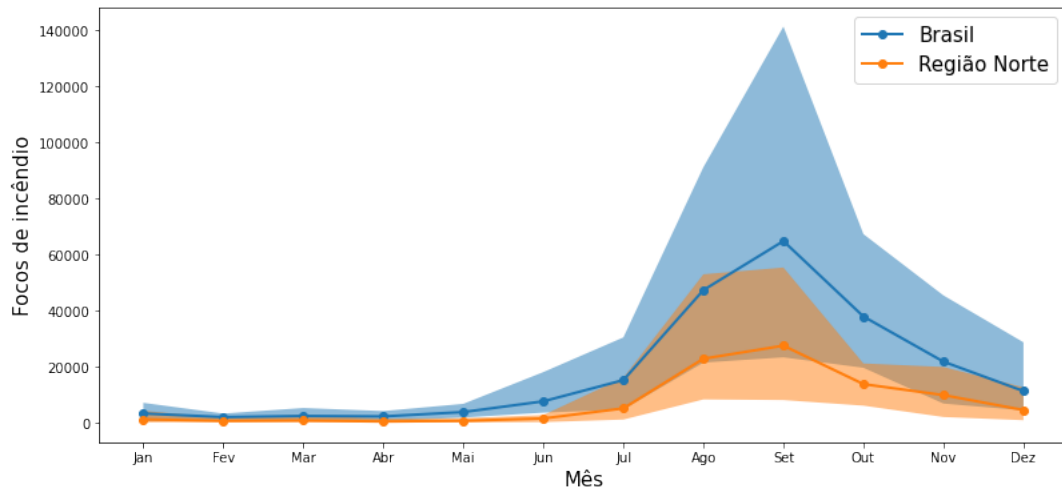
### 3.4.2 Queimadas

A Figura 3 mostra o comparativo mensal do número de focos de incêndio para a série histórica de 1999 a 2022 para o Brasil e a região Norte. A Figura representa as médias (linhas contínuas) e os máximos e mínimos (região colorida em torno da média). Neste gráfico percebe-se a forte correlação entre a estação das chuvas no Norte do Brasil e o número de focos de incêndio, que se concentram entre os meses de julho a dezembro, destacando-se os meses de agosto e setembro com os maiores registros de focos de incêndio.

A Figura 4 traz o comparativo anual do período 1999 a 2022 para o Brasil, considerando-se todas as regiões, e para a região Norte isoladamente. De maneira geral, um crescimento ou queda no número de focos de incêndio registrados de um ano para outro no Brasil também ocorre na região Norte, como ocorre entre os anos de 2000 e 2002 e entre 2013 e 2015, por exemplo.

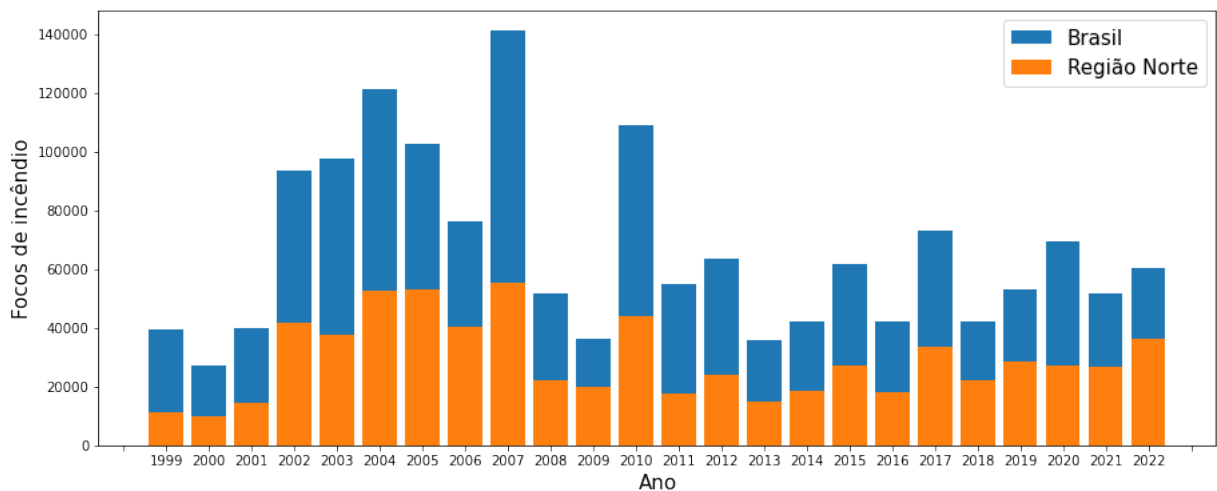
Um recorte por bioma dos dados anuais de incêndios é apresentado na Figura 5. Percebe-se que o Bioma Amazônia lidera o número de focos de incêndio em toda a série histórica. Comparando com a série histórica anual da Figura 4, nota-se que dos 3 anos com os maiores números de focos de incêndio registrados, no ano de 2004 foram as queimadas registradas no Bioma Amazônia que tiveram recorde de registros, enquanto nos anos de 2007 e 2010, foi a combinação de elevados números de focos de incêndio registrados no Bioma Amazônia e no Bioma Cerrado que elevaram os números nacionais. Além disso,

Figura 3 – Comparativo mensal Brasil e região Norte.



Fonte: Elaborado pela autora com base nos dados do BDQUEIMADAS (INPE, c).

Figura 4 – Comparativo anual Brasil e região Norte.



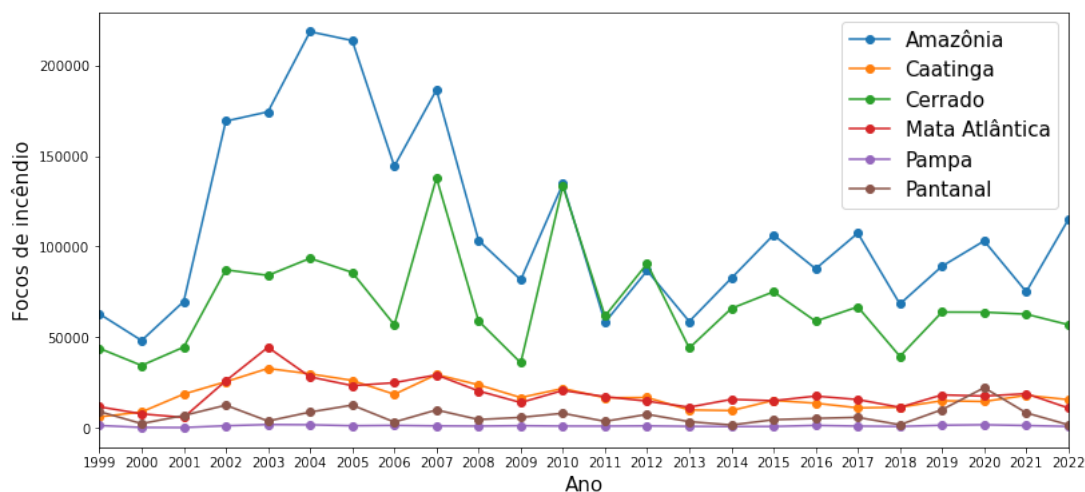
Fonte: Elaborado pela autora com base nos dados do BDQUEIMADAS (INPE, c).

comparando com os dados da Figura 2, o ano de 2004 registrou não apenas recordes de focos de incêndios, mas também de área desmatada.

Analisando-se a série histórica do número de focos de incêndio registrados nos nove estados que compreendem o Bioma Amazônia, Figura 6, conclui-se que nos anos com maiores números de focos de incêndio também houve recorde nos registros de incêndio nos estados do Mato Grosso e Pará, que são os estados onde, em geral, mais se registrou focos de incêndios ao longo dos anos contemplados pelos dados do INPE.

De acordo com Volpato *et al.* (2022), a tendência de redução nos focos de incêndio

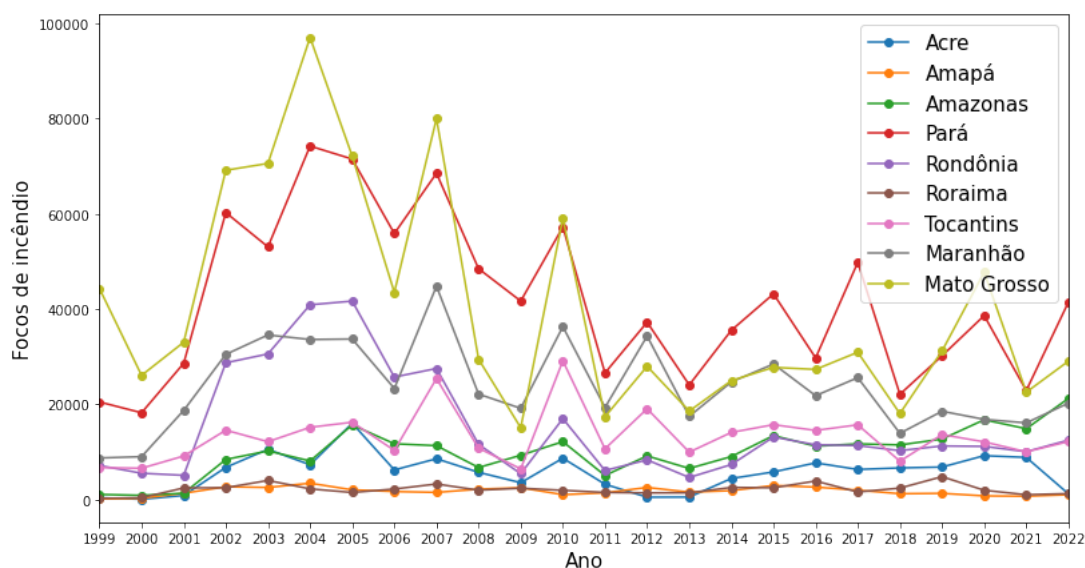
Figura 5 – Número de focos de incêndio por Bioma por ano.



Fonte: Elaborado pela autora com base nos dados do BDQUEIMADAS (INPE, c).

no Mato Grosso a partir do ano 2005 pode estar relacionada à adoção do sistema de plantio direto de culturas temporárias pelos grandes agricultores, que substitui a agricultura convencional de corte e queima dos restos, assim como à Lei Complementar nº 233, de 21 de dezembro de 2005, que proibiu o uso do fogo para limpeza e manejo de áreas na estação seca. De acordo com os mesmo autores, os extremos de 2007 e 2010 podem ser explicados pelas condições climáticas observadas nestes anos (acúmulo de biomassa vegetal devido aos anos anteriores terem sido mais chuvosos).

Figura 6 – Número de focos de incêndio por Estado do Bioma Amazônia.



Fonte: Elaborado pela autora com base nos dados do BDQUEIMADAS (INPE, c).

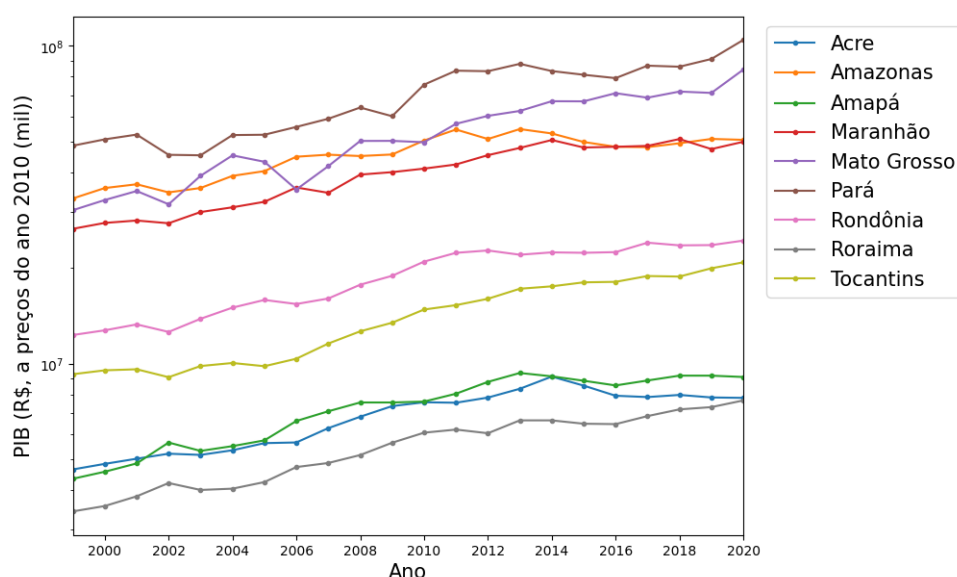


### 3.4.3 PIB

O Produto Interno Bruto (PIB) é a soma de todos os bens e serviços finais produzidos por um país, estado ou cidade, nos preços em que chegam ao consumidor (IBGE, 2023a). O IBGE calcula o PIB com dados produzidos pelo próprio órgão e alguns dados externos e divulga o resultado do PIB de cada ano no 2º trimestre do ano seguinte.

A Figura 7 apresenta o PIB dos nove estados que compõem a região amazônica entre os anos de 1999 e 2020. Os estados com maior PIB da região são Pará, Mato Grosso, Amazonas e Maranhão. Os dois primeiros estados são também os que apresentaram maiores taxas de desmatamento e registros de focos de incêndio na série histórica. Possivelmente, essas variáveis estão relacionadas devido à pressão ambiental em função da maior atividade econômica, incluindo a expansão das fronteiras agropecuárias.

Figura 7 – PIB dos estados da região amazônica em escala Logarítmica.



Fonte: Elaborado pela autora com base nos dados do Ipeadata (IPEA, 2023).

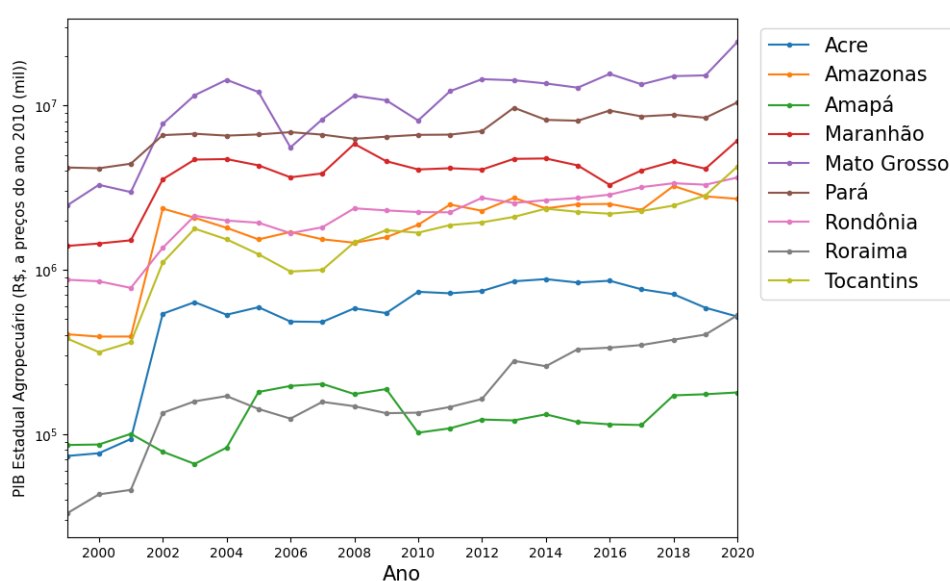
### 3.4.4 PIB Estadual Agropecuário

O IBGE calcula anualmente o valor bruto da produção agropecuária por município, em todo o território nacional. O PIB Estadual Agropecuário compreende a agricultura, pecuária, silvicultura e exploração florestal e pesca (IBGE, 2016).

A Figura 8 apresenta os dados do IBGE entre os anos 1999 e 2020 para os estados da região amazônica. Com exceção do Amapá, observa-se um aumento acentuado no PIB dos demais estados a partir do ano 2002. Isto se deve, principalmente, à mudanças na metodologia de cálculo, cujos detalhes não serão abordados neste trabalho, mas encontram-se descritos no relatório metodológico IBGE (2008).

Os estados com maior PIB Agropecuário da região são o Mato Grosso, Pará e Maranhão. Em 2020 foi registrado o maior valor desta série histórica nestes 3 estados e também em Tocantins, Rondônia e Roraima. Neste ano, o PIB Agropecuário brasileiro registrou recorde, chegando a quase 27% do PIB, reflexo dos preços maiores na comparação com 2019 e uma safra recorde de grãos, além dos maiores preços das proteínas frente a 2019 e a expansão da produção e abate de suínos e aves e da produção de ovos e leite (CNA, 2023).

Figura 8 – PIB Estadual Agropecuário dos estados da região amazônica em escala Logarítmica.



Fonte: Elaborado pela autora com base nos dados do Ipeadata (IPEA, 2023).

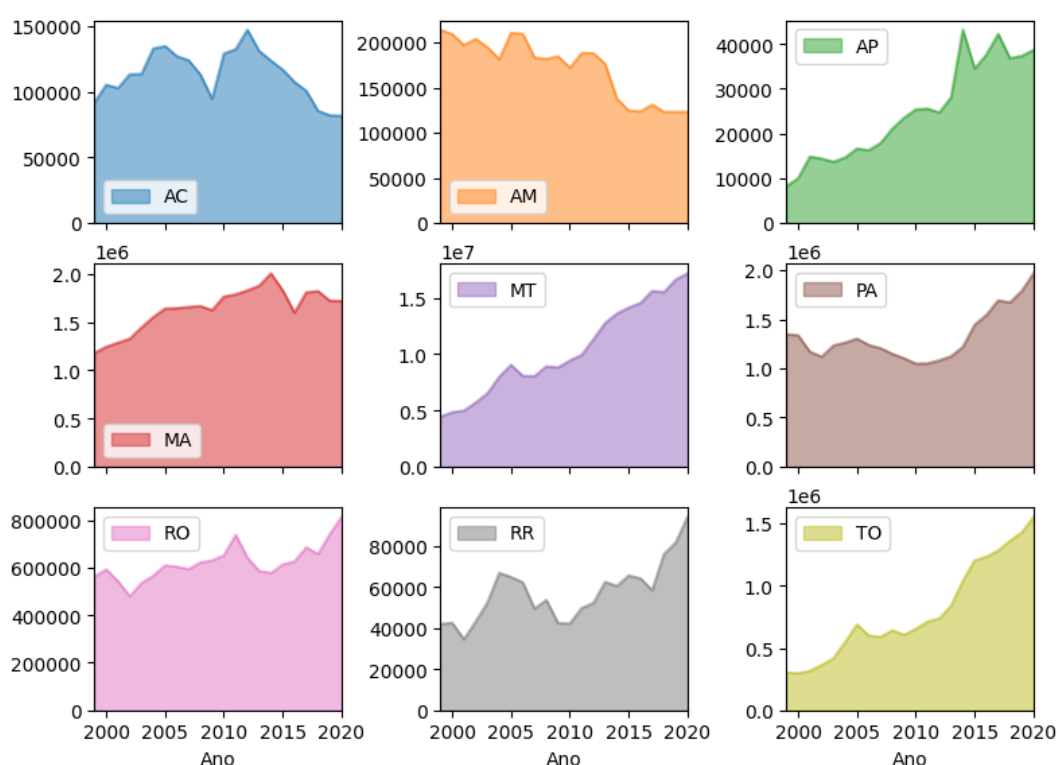
### 3.4.5 Área plantada ou destinada à colheita

O IBGE realiza o levantamento sistemático da produção agrícola, que consiste em uma pesquisa de previsão e acompanhamento da safra agrícola das 35 principais culturas cultivadas no país. O levantamento é mensal, em todas as unidades da federação, com acompanhamento desde a fase de intenção de plantio até a conclusão da colheita (IBGE, 2002).

De acordo com a Figura 9, os estados do Acre e Amazonas sofreram uma redução na área plantada desde 2012, atingindo a menor área plantada em 2020. Estes dois estados foram os únicos a apresentar retração na área plantada ou destinada à colheita entre os anos de 1999 e 2020. O Amapá e o Tocantins apresentaram tendência de crescimento ao longo dos anos analisados, tendo quintuplicado a área plantada entre 1999 e 2020, assim também como o Mato Grosso, que praticamente quadruplicou a área plantada neste período.

O Mato Grosso é o estado com maior área plantada ou destinada à colheita ao longo de toda a série histórica. Em 2020, o Mato Grosso tinha uma área plantada quase nove vezes maior do que o Pará, segundo estado com maior área plantada da região. O Pará apresentou crescimento desde 2010, tendo quase que dobrado a área plantada em dez anos. Mato Grosso, Pará, Tocantins, Rondônia e Roraima tinham em 2020 as maiores áreas plantadas da série histórica.

Figura 9 – Área plantada ou destinada à colheita (Ha).



Fonte: Elaborado pela autora com base nos dados do Ipeadata (IPEA, 2023).

### 3.4.6 Valor da produção agropecuária

O IBGE calcula o valor da produção agropecuária multiplicando a quantidade de cada produto obtida na área colhida pelo preço médio ponderado e o preço médio pago ao produtor por animal ou produção animal. A pesquisa é anual, por município. São considerados 29 produtos agrícolas de culturas temporárias e 33 de culturas permanentes, e na pecuária, além dos efetivos das espécies animais criadas, também considera-se as produções de leite, lã, ovos de galinhas e de codornas, mel e casulos de bicho-da-seda. O efetivo da pecuária compreende bovinos, bubalinos, equinos, asininos, muares, suínos, ovinos, caprinos, coelhos, codornas, galinhas e o número total dos demais galináceos (galos, frangas, frangos e pintos) (IBGE, 2002).

O valor da produção agropecuária é apresentado na Figura 10. Seguindo o comportamento das variáveis anteriores, os estados do Mato Grosso e Pará possuem maior valor de produção agropecuária, o que é esperado tendo em vista a maior área plantada e maiores áreas desmatadas da região amazônica.

O ano de 2020 registrou o maior valor de produção ao longo desta série histórica nos estados do Mato Grosso, Pará, Tocantins, Maranhão, Rondônia e Roraima. O aumento se deve aos recordes nas safras de grãos nos anos 2019 e 2020 (Agência IBGE Notícias, 2022). O estado com maior valor de produção agropecuária é o Mato Grosso, cujo valor em 2020 foi cinco vezes maior do que o do Pará. Acre e Amapá apresentaram tendência de queda desde 2014.

O valor da produção agropecuária no Mato Grosso aumentou mais de seis vezes ao longo dessa série histórica e o estado do Tocantins foi o que teve o maior aumento da série, multiplicando em mais de nove vezes o valor da produção entre 1999 e 2020. O único estado que apresentou retração considerando-se os anos iniciais e finais da série foi o Acre.

Em todos os estados houve um aumento expressivo no valor da produção no ano de 2011, após alguns anos de baixa. Este aumento foi impulsionado, de maneira geral, pela elevação dos preços dos produtos agrícolas tanto no mercado interno quanto externo. Entre os principais produtos responsáveis pelo aumento naquele ano, estão a soja, a cana-de-açúcar e o milho, que sozinhos respondiam a 57,2% do valor da produção brasileira (Agência IBGE Notícias, 2012).

#### 3.4.7 População

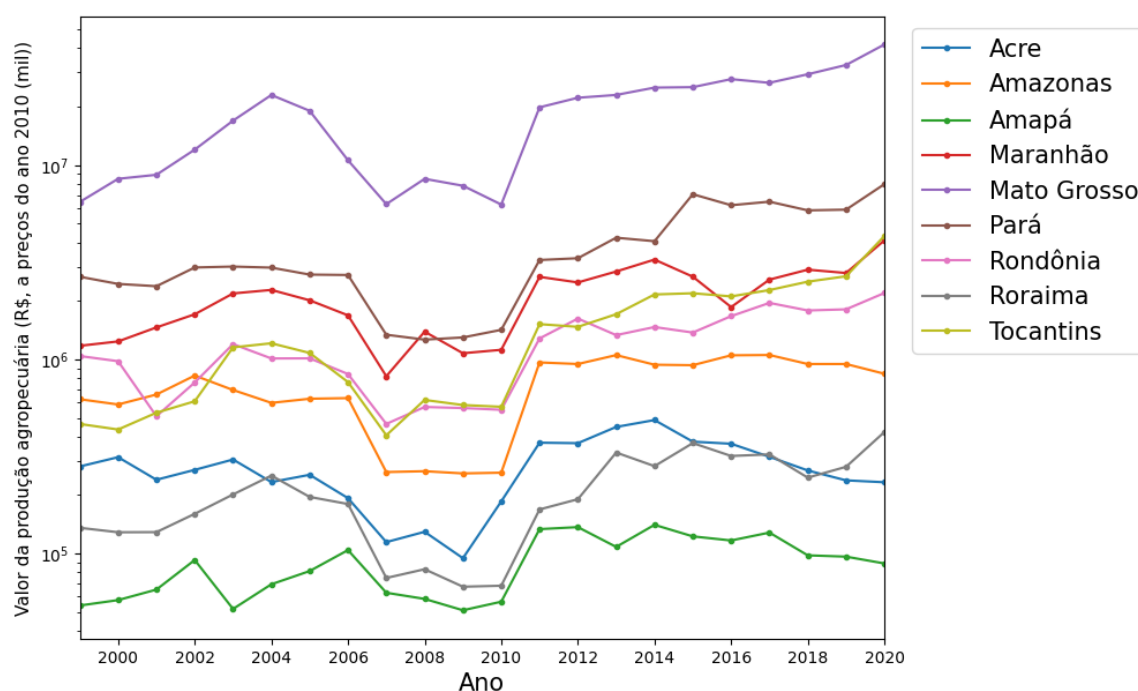
O IBGE realiza censos demográficos a cada dez anos, desde 1940. O único adiamento no Censo ocorreu em 2020 devido à pandemia da Doença por Coronavírus 2019 - COVID-19, que só foi realizado em 2022. Os dados populacionais para os anos em que não são realizados Censos são estimativas obtidas através de uma metodologia na qual se observa a tendência de crescimento populacional, entre dois censos demográficos consecutivos, em relação à tendência de crescimento de uma área geográfica maior (IBGE, 2021).

Na região amazônica brasileira, os estados mais populosos são Pará, Maranhão e Amazonas, com aproximadamente nove, sete e quatro milhões de habitantes em 2020, respectivamente, como mostra a Figura 11. Os menos populosos são Roraima, Acre e Amapá, cujas populações em 2020 não chegavam a 1 milhão de habitantes. Ao longo da série, o crescimento populacional mais acentuado ocorreu no Pará, seguido de Maranhão e Amazonas.

#### 3.4.8 Fluxo de crédito rural

O fluxo de crédito rural são as concessões de crédito rural (empréstimos ou financiamentos) realizadas no âmbito do Sistema Financeiro Nacional, para aplicação exclusiva

Figura 10 – Valor da produção agropecuária dos estados da região amazônica em escala Logarítmica.



Fonte: Elaborado pela autora com base nos dados do Ipeadata (IPEA, 2023).

nas finalidades e condições estabelecidas pelo BCB (Banco Central do Brasil, 2023).

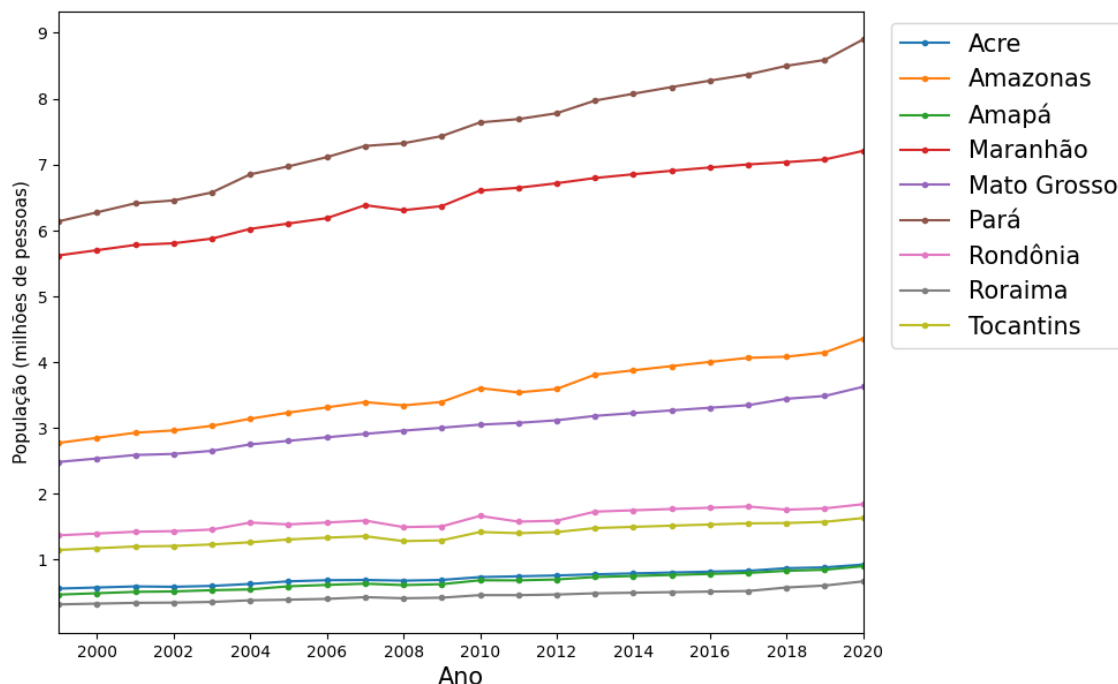
Na Figura 12 são apresentados os dados do BCB para os estados que compõem a região amazônica. No estado do Mato Grosso concentram-se os maiores valores de concessões de crédito rural em todos os anos da série. No ano de 2012, o total em crédito concedido atingiu valor máximo no Mato Grosso, Tocantins, Maranhão, Amazonas e Acre. O mesmo ocorreu em 2020 nos estados do Pará, Rondônia e Roraima. O aumento mais acentuado no fluxo de crédito ao longo da série ocorreu no Mato Grosso, Tocantins e Maranhão.

### 3.4.9 Extensão de UCs federais e estaduais

O Instituto Socio Ambiental (ISA), através do Programa de Monitoramento de Áreas Protegidas, realiza o acompanhamento da criação, revogação e alteração de limites e categorias de Unidades de Conservação (UCs) e atualiza diariamente as informações na sua página da internet (Instituto SocioAmbiental, 2023). Existem atualmente 148 municípios com incidência de UCs de Proteção Integral e 282 com UCs de Uso Sustentável na região amazônica (Murer, B. M and Futada, S. M., 2023).

A primeira unidade de conservação federal na amazônia foi criada em 1959, no estado do Acre, e a estadual em 1978, no estado do Mato Grosso. A Figura 13 mostra a

Figura 11 – População dos estados da região amazônica.



Fonte: Elaborado pela autora com base nos dados do Ipeadata (IPEA, 2023).

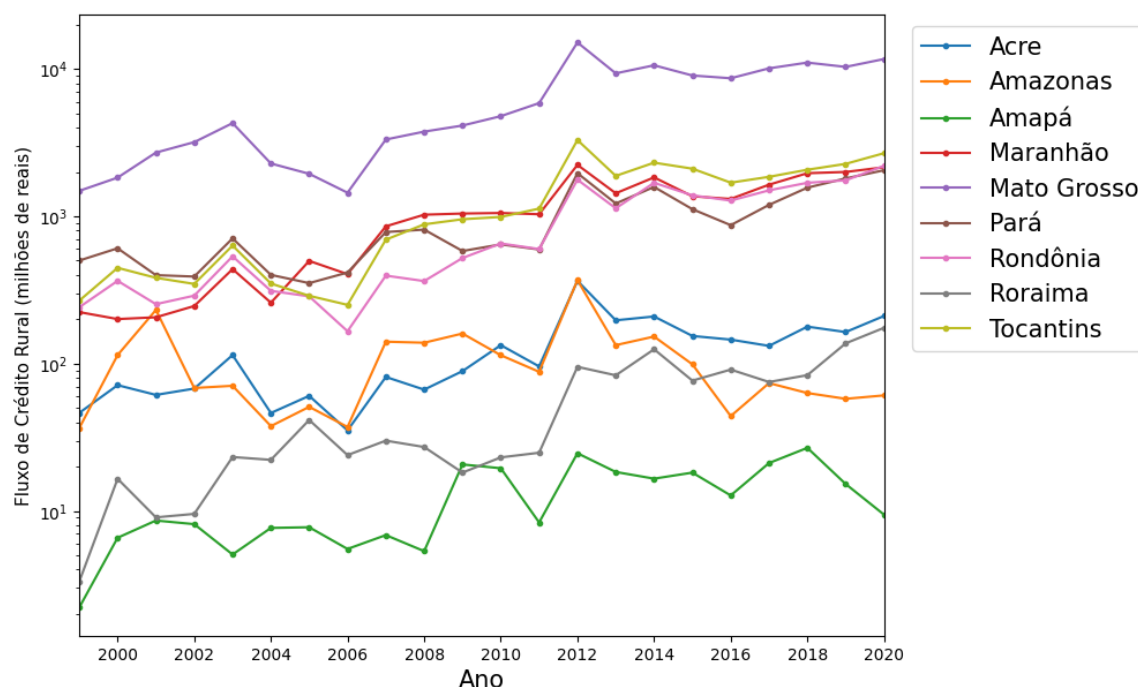
soma acumulada da área das unidades de conservação federais e estaduais por estado da Amazônia.

O estado do Amazonas possuía a maior área total e o maior número de unidades de conservação em 2020, sendo mais de 43 milhões de hectares em 81 unidades, isso representa quase 60% de seu território (Murer, B. M and Futada, S. M., 2023). Em seguida vem o estado do Pará, com mais de 31 milhões de hectares e 75 unidades de conservação em 2020. Na sequência estão os estados do Amapá, Rondônia, Acre, Mato Grosso, Roraima, Maranhão e Tocantins, estes dois últimos com pouco mais de 1,5 milhões de hectares em 2020.

Considerando a série histórica, o período de 1999 a 2009 foi o mais intenso em criação de unidades de conservação federais na região amazônica, sendo que 90% das UCs federais da série analisada foram criadas neste período, a maior parte nos estados do Amazonas e Pará. As demais foram criadas entre 2014 e 2018.

O estado com maior área destinada à UCs estaduais é o Pará, que sozinho tem uma área igual à aproximadamente a soma das áreas de UCs estaduais dos demais estados da região. Maior parte desta área foi criada no ano de 2006.

Figura 12 – Fluxo de crédito rural dos estados da região amazônica em escala Logarítmica.



Fonte: Elaborado pela autora com base nos dados do Ipeadata (IPEA, 2023).

#### 3.4.10 Efetivo Bovino

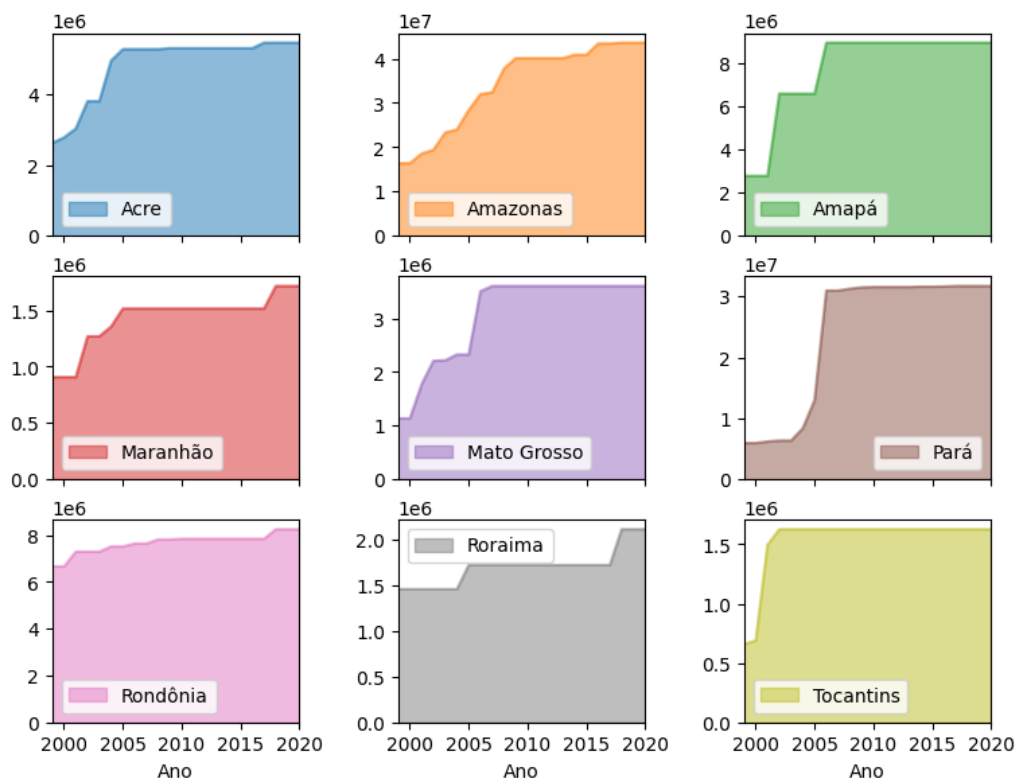
O IBGE realiza o levantamento do número de animais através do Censo Agropecuário desde o início do século 20. O efetivo bovino “compreende o estoque total de animais vivos existentes no município, na data de referência da pesquisa, independente de raça ou finalidade da criação (corte, leite, trabalho)” (IBGE, 1998).

Os dados do IBGE para os estados do bioma Amazônia estão representados na Figura 14. Os maiores efetivos bovinos concentram-se nos estados do Mato Grosso, Pará e Rondônia. Com exceção do Amapá, todos os estados apresentam tendência de crescimento ao longo da série. O Amapá registrou tendência de queda no número de cabeças de gado entre os anos de 2015 e 2020.

#### 3.4.11 Área colhida de lavoura permanente e temporária

Anualmente, é realizado, dentro do Censo Agropecuário, o levantamento da área colhida das lavouras permanentes e temporárias (IBGE, 1998). Culturas temporárias possuem curta ou média duração e após a colheita necessitam de novo plantio para produzir. Já as culturas permanentes possuem longo ciclo vegetativo, que permitem colheitas sucessivas, sem necessidade de novo plantio. São exemplos de lavoura permanente: açaí, banana, borracha, cacau e café, e de lavoura temporária: algodão, arroz, cana-de-

Figura 13 – Extensão de UCs (Ha, total acumulado).



Fonte: Elaborado pela autora com base nos dados do ISA (Instituto SocioAmbiental, 2023).

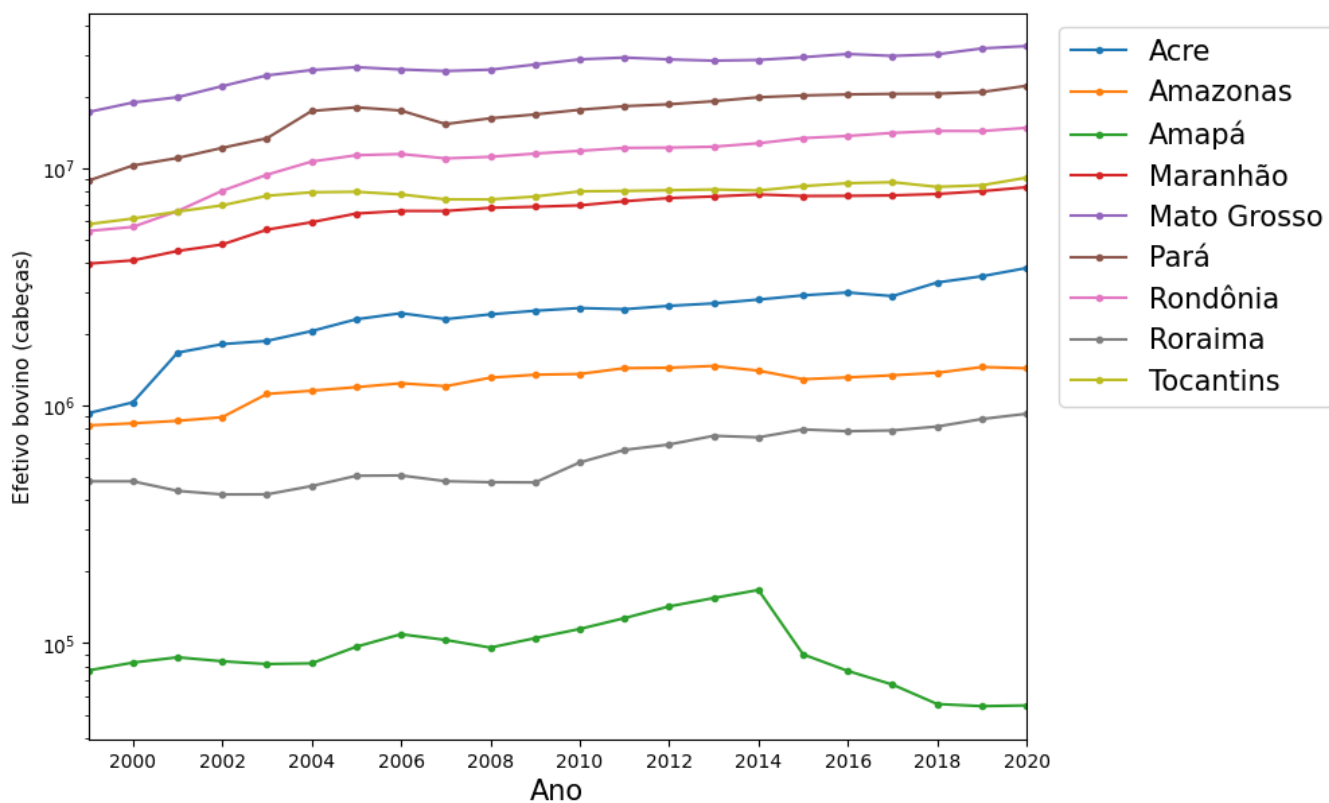
açúcar, milho e soja (IBGE Educa, 2023).

As Figuras 15 e 16 apresentam os dados por estado referente às áreas colhidas das lavouras permanentes e temporárias, respectivamente. Observa-se uma tendência de queda nas áreas colhidas de lavoura permanente nos estados do Amazonas, Mato Grosso e Rondônia ao longo da série. No estado do Maranhão a tendência de queda é observada entre os anos de 2008 e 2020. O estado do Pará apresentou um crescimento acentuado na área colhida de culturas permanentes nos últimos anos, ultrapassando o estado do Mato Grosso em 2016 e registrando a maior área colhida da região no ano de 2020, sendo este o recorde desta série histórica.

No que tange à área colhida de lavoura temporária, a tendência é oposta nos estados do Mato Grosso e Rondônia, apresentando crescimento, que também é verificado no Amapá, Roraima, Tocantins e Pará. Apenas o Amazonas indica uma tendência de queda na área colhida das culturas temporárias ao longo da série, assim como o Acre a partir de 2014. O Mato Grosso registrou o recorde da região no ano de 2020 para esta série histórica.



Figura 14 – Efetivo Bovino em escala Logarítmica (cabeça).



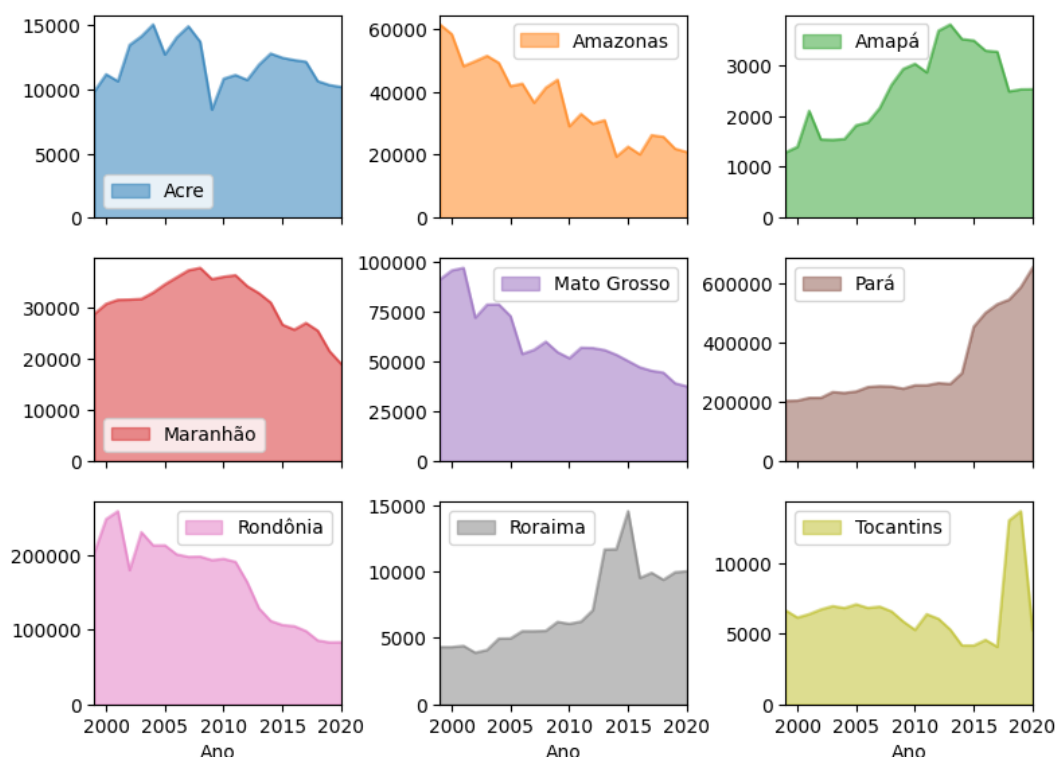
Fonte: Elaborado pela autora com base nos dados do Ipeadata (IPEA, 2023).

### 3.4.12 Produção de Milho e Soja

No Censo Agropecuário do IBGE, são registrados os valores das produções dos principais cultivos. Para o milho, “considera-se o beneficiamento de milho em estabelecimento agrícola e a produção de sementes de milho quando atividades complementares ao cultivo, não considerando o cultivo de milho forrageiro, a produção de sementes certificadas de milho, inclusive modificadas geneticamente, os serviços de preparação de terreno, cultivo e colheita realizados sob contrato, o beneficiamento de milho realizado em estabelecimento agrícola sob contrato, a moagem e a fabricação de farinha de milho e derivados” (IBGE, 1998). Para a soja, “considera-se a produção de semente de soja quando atividade complementar ao cultivo, não considerando a produção de sementes certificadas de soja, inclusive modificadas geneticamente, os serviços de preparação de terreno, cultivo e colheita realizados sob contrato, a produção de óleo de soja em bruto, inclusive torta e farelo, o refino de óleo de soja, a fabricação de margarinas e gorduras de soja e a produção de biocombustível” (IBGE, 1998).

As Figuras 17 e 18 apresentam os valores das produções de milho e soja, respectivamente. Os estados do Mato Grosso, Maranhão e Pará possuem maior produção de milho

Figura 15 – Área colhida de lavoura permanente (Ha).



Fonte: Elaborado pela autora com base nos dados do Ipeadata (IPEA, 2023).

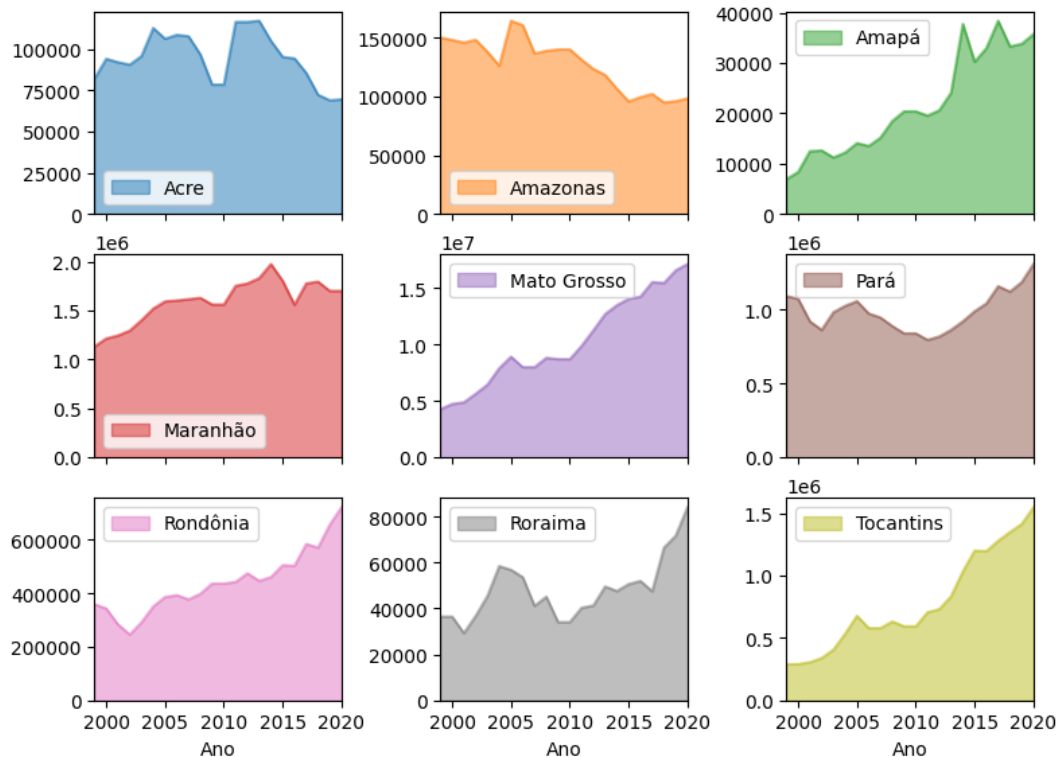
na região. Mato Grosso, Maranhão e Tocantins são os maiores produtores de soja.

### 3.4.13 Despesa orçamentária por função - gestão ambiental, agricultura e organização agrária

A despesa orçamentária é toda transação que depende de autorização legislativa para ser efetivada e compreende todos os gastos realizados pelos entes públicos. Ela é classificada de acordo com funções e subfunções, que agregam os gastos públicos por área de ação governamental nas três esferas de Governo. A despesa aqui considerada é a soma das despesas do poder público com as funções Gestão Ambiental, Agricultura e Organização Agrária (MCASP, 2018).

A Figura 19 mostra a despesa orçamentária por ano e por estado da região Amazônica. De forma geral, o valor gasto com gestão ambiental, agricultura e organização agrária pelo poder público tende a crescer ao longo da série. De 2014 a 2020, os maiores gastos com esta função orçamentária se concentraram nos estados de Mato Grosso e Pará.

Figura 16 – Área colhida de lavoura temporária (Ha).



Fonte: Elaborado pela autora com base nos dados do Ipeadata (IPEA, 2023).

#### 3.4.14 Partido do Governador em Exercício

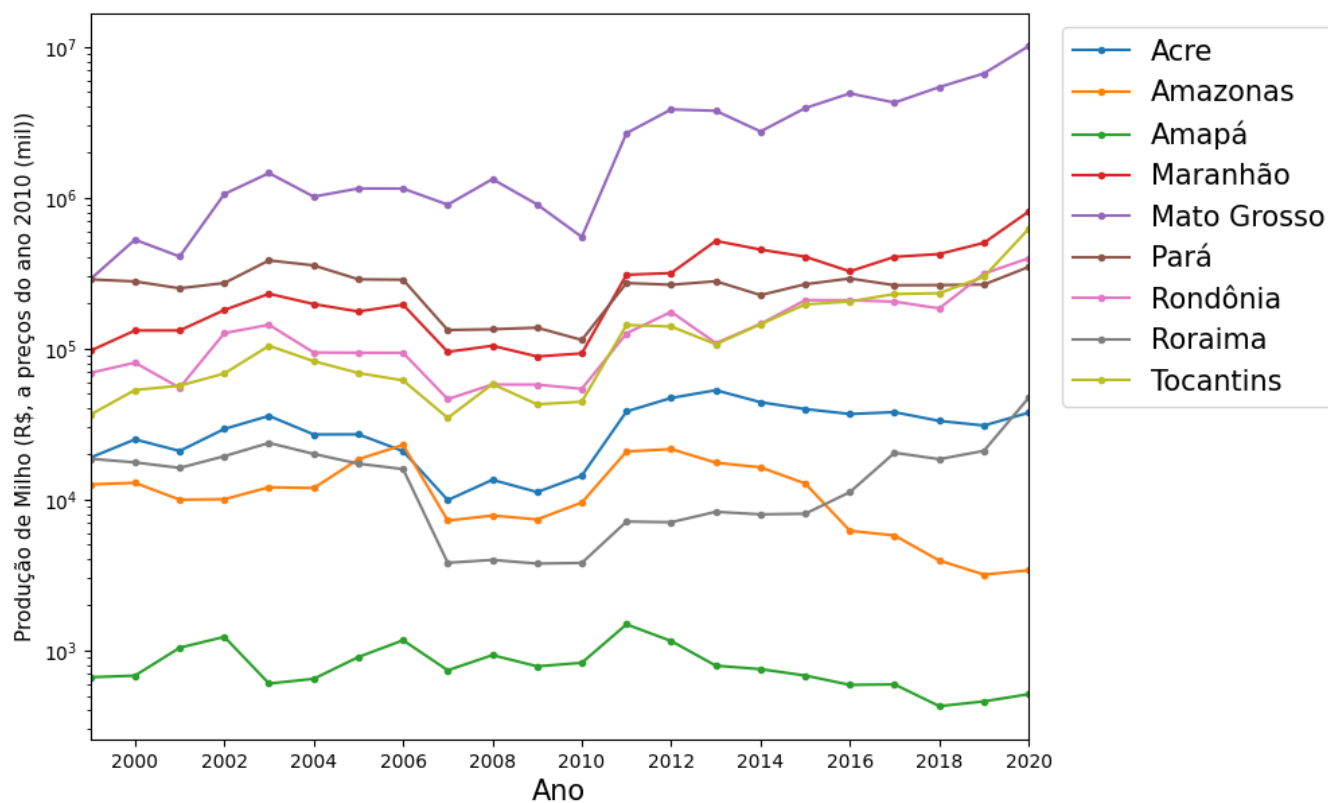
Os partidos dos governadores em exercício entre os anos de 1999 e 2020 de cada um dos nove estados que compõem a região amazônica foram obtidos da Wikipedia (WIKIPÉDIA, 2023c; WIKIPÉDIA, 2023e; WIKIPÉDIA, 2023d; WIKIPÉDIA, 2023f; WIKIPÉDIA, 2023g; WIKIPÉDIA, 2023h; WIKIPÉDIA, 2023a; WIKIPÉDIA, 2023b; WIKIPÉDIA, 2023i). A intenção é verificar se a sigla do partido tem importância para a variável alvo (desmatamento).

#### 3.4.15 Matriz de correlação

A matriz de correlação pode ser vista na Figura 20. Para melhor visualização, os partidos políticos foram omitidos para esta análise. As maiores correlações encontradas foram entre a variável “Produção de soja” e as variáveis “Produção de milho” (coeficiente de correlação 0,92), “Área plantada ou destinada” (0,85) e “Valor da produção agropecuária” (0,81). A variável “Efetivo bovino” também apresentou forte correlação com a variável “Área Colhida de Lavoura Temporária” (0,87), assim como a variável “Produção de milho” com a variável “Valor da produção agropecuária” (0,82).

Tomando apenas o período 2000 a 2006, Rivero *et al.* (2009), considerando uma

Figura 17 – Produção de Milho (R\$, a preços do ano 2010 (mil)) em escala Logarítmica.



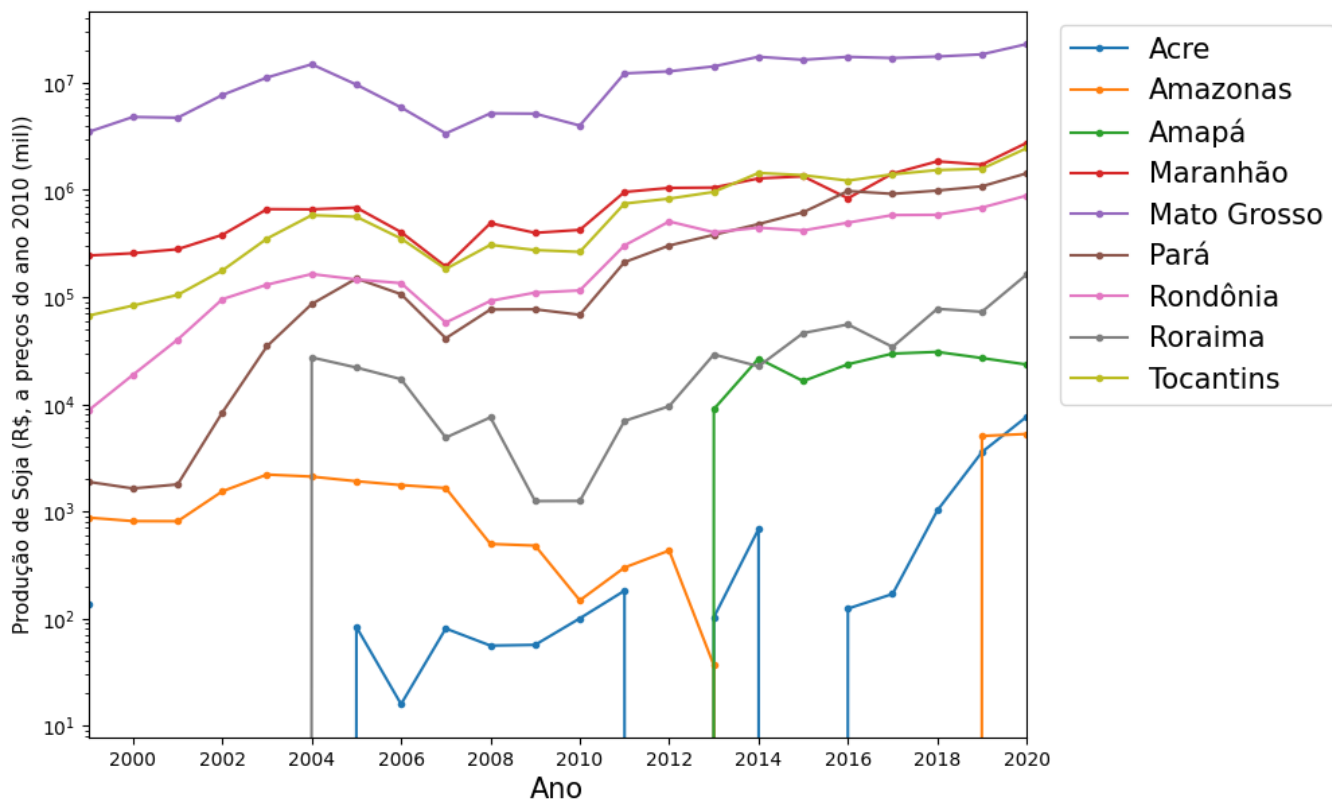
Fonte: Elaborado pela autora com base nos dados do Ipeadata (IPEA, 2023).

seleção de 782 municípios da Amazônia, encontraram um coeficiente de correlação de 0,73 entre o número de cabeças de gado e o desmatamento. No presente trabalho, considerando-se toda a região Norte do país e os estados de Mato Grosso e Maranhão, as maiores correlações com o desmatamento ocorreram com a variável “Área Colhida de Lavoura Permanente”, com coeficiente igual a 0,51, “Efetivo Bovino”, com coeficiente 0,47, “Área Colhida de Lavoura Temporária” com coeficiente 0,42, “PIB Agropecuário”, com coeficiente igual a 0,36 e com a variável “Focos de incêndio”, com coeficiente 0,24.

Correlações negativas ocorreram entre a variável “Extensão de UCs” e as variáveis “Produção de soja”, “Produção de milho”, “Efetivo bovino”, “Fluxo de crédito rural”, “Valor da produção agropecuária” e “Área plantada ou destinada”. Também possuem correlação negativa “Fluxo de crédito rural” e “Área colhida de lavoura permanente”, assim como a variável “Área colhida de lavoura permanente” com as variáveis “Produção de milho” e “Produção de soja”.

A matriz de correlação considerando a variável alvo “Desmatamento” e os partidos políticos possui correlação negativa para todos os partidos, exceto “PSDB” com coeficiente de correlação 0,38, “PPS”, com coeficiente 0,22, “MDB”, com 0,20 e “PSC”, com 0,025.

Figura 18 – Produção de Soja (R\$, a preços do ano 2010 (mil)) em escala Logarítmica.

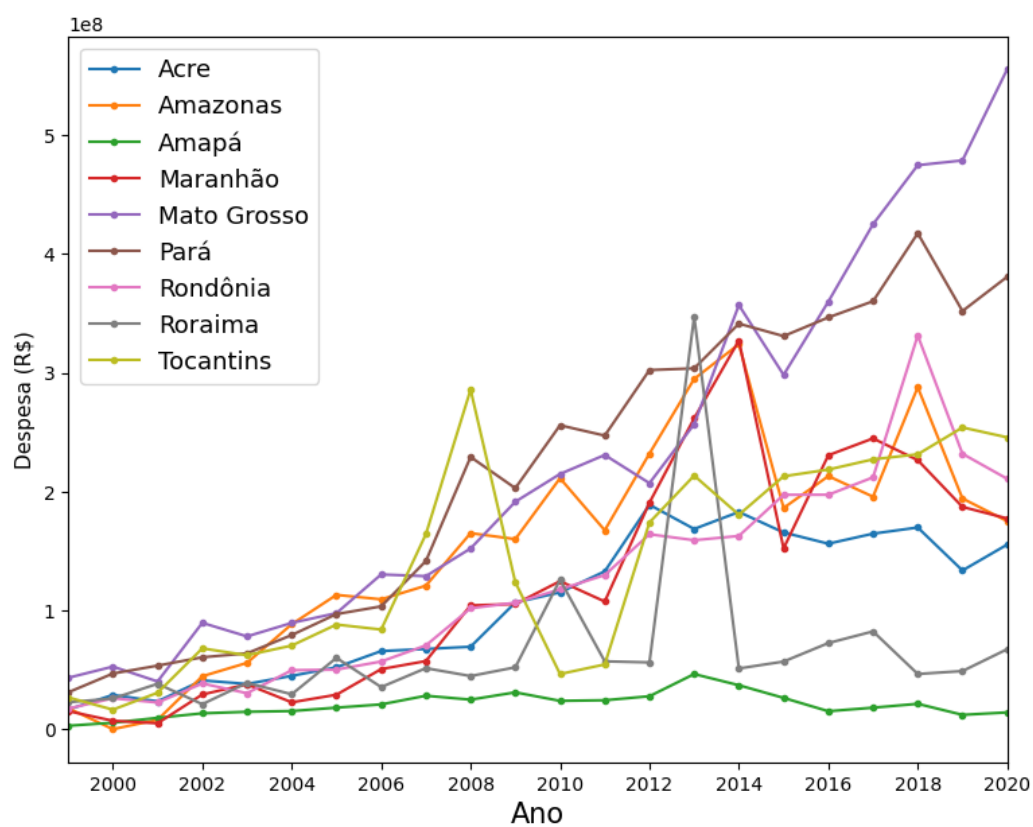


Fonte: Elaborado pela autora com base nos dados do Ipeadata (IPEA, 2023).

Com relação às demais variáveis, quando correlacionadas aos partidos políticos, todos os coeficientes ficam abaixo de 0,4, exceto a variável “Área colhida de lavoura permanente” e “MDB”, com coeficiente 0,49.

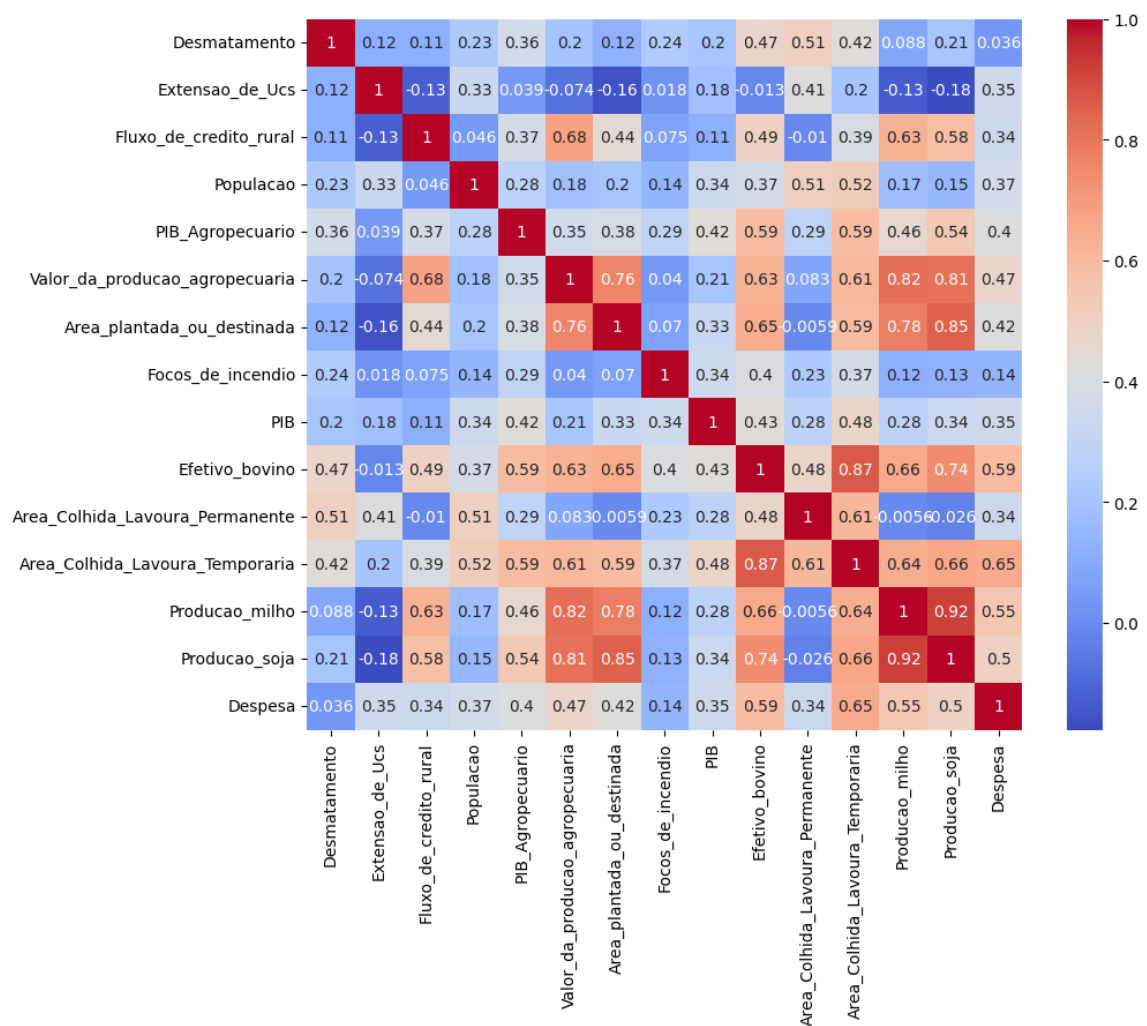
Apesar de algumas variáveis apresentarem correlação alta, optou-se por manter todas as variáveis, já que o número total não é tão alto (16) e não impacta a execução dos algoritmos.

Figura 19 – Despesa orçamentária por função - gestão ambiental, agricultura e organização agrária.



Fonte: Elaborado pela autora com base nos dados do Ipeadata (IPEA, 2023).

Figura 20 – Matriz de correlação das variáveis quantitativas.



Fonte: Elaborado pela autora.





## 4 MÉTODOS

Neste capítulo são apresentados os modelos de regressão escolhidos para abordar o problema proposto. Cinco modelos foram selecionados, quatro modelos de árvore: Árvore de Decisão, Floresta Aleatória, Extra Trees e Gradient Boosting e o modelo Máquinas de Vetores-Suporte. Em seguida são apresentadas as medidas de desempenho utilizadas para avaliar os modelos: coeficiente de determinação, erro quadrático médio e erro absoluto médio.

### 4.1 Algoritmos

#### 4.1.1 Árvores de Decisão

Este algoritmo, para problemas de regressão, é também chamado de Árvores de regressão. Nas Árvores de Decisão (AD), o conjunto de dados é particionado em grupos menores (nós) e um modelo simples é ajustado para cada subgrupo. O resultado final é obtido da sequência de todas as decisões desde o nó raiz (nó inicial) até o nó terminal ou folha, no qual não há mais ramificações e onde encontram-se as decisões para uma determinada trajetória seguida na árvore. Os nós terminais da árvore contêm valores numéricos previstos, que são a média dos valores alvo para dada folha. Em cada nó da árvore é localizado um atributo de onde partem as ramificações com os valores possíveis para cada um. Portanto, testa-se cada atributo cobrindo-se todas as combinações possíveis em uma estrutura hierárquica análoga a um fluxograma. As árvores permitem a visualização das etapas seguidas em cada decisão, permitindo que o usuário considere diferentes cenários e suas consequências (MITCHELL, 1997).

##### 4.1.1.1 Floresta Aleatória e Extra Trees

O algoritmo Floresta Aleatória (FA) combina várias árvores de decisão, cada uma sendo gerada usando-se um subconjunto aleatório dos atributos (ROMERO, 2022). Para selecionar os exemplos de treinamento, o algoritmo utiliza *bootstrap*, que consiste em criar um novo conjunto de dados com o mesmo número de elementos do conjunto original, porém com elementos escolhidos aleatoriamente sendo que alguns são repetidos e outros podem estar ausentes. Podem ser gerados assim vários conjuntos de dados com o mesmo tamanho a partir da amostragem do conjunto original. Definidas as amostras, utiliza-se a técnica *bootstrap aggregation* ou *bagging*, na qual as amostras *bootstrap* geradas são utilizadas para treinar o modelo preditivo (BISHOP, 2006). A previsão final será a média das previsões de todas as árvores. Essas modificações levam as florestas aleatórias a possuírem uma capacidade preditiva superior à de uma árvore isolada.

No modelo de árvores extremamente aleatórias, do inglês, Extra Trees (ET), assim como no algoritmo de florestas aleatórias, muitas árvores de decisão são geradas, mas a amostragem para cada árvore é aleatória e sem reposição. As variáveis também são selecionadas aleatoriamente para cada árvore assim como a divisão dos nós.

#### 4.1.2 Gradient Boosting

Assim como no algoritmo Floresta Aleatória, o Gradient Boosting (GB) também combina várias árvores com objetivo melhorar a acurácia. A técnica de boosting é uma técnica de aprimoramento, baseada na regra do gradiente estocástico, que visa minimizar o erro cometido pela árvore anterior, caminhando em direção oposta ao gradiente da função erro. A ideia é treinar o conjunto de dados com um modelo preditivo e testá-lo obtendo-se a medida do erro cometido. No próximo treinamento do modelo, os erros cometidos serão minimizados e o conjunto testado novamente e assim sucessivamente, até que atinja-se o erro desejado (ROMERO, 2022). Ou seja, diferentemente do modelo FA, as árvores são construídas e treinadas em sequência e cada uma usa uma forma ponderada do conjunto de dados em que o coeficiente de ponderação depende do desempenho das árvores anteriores (BISHOP, 2006).

#### 4.1.3 Máquinas de vetores-suporte

Nos problemas de regressão, o objetivo do modelo SVM é encontrar uma função que aproxima bem os dados de amostra. Para isso, primeiro é preciso encontrar um preditor que apresente o menor erro dentre um conjunto de preditores. Para avaliar o desempenho de um preditor, utiliza-se uma função perda que permite determinar como os erros serão penalizados em cada ponto da amostra (VERDÉRIO, 2015).

### 4.2 Medidas de desempenho

A seguir são descritas as métricas utilizadas para avaliar o desempenho dos algoritmos acima citados. Essas medidas são tradicionalmente utilizadas em modelos de regressão.

#### 4.2.1 Coeficiente de determinação ( $R^2$ )

O coeficiente de determinação ( $R^2$ ) é utilizado para verificar a qualidade do ajuste do modelo. É definido como a proporção da variação da variável resposta  $Y$  explicada pelo modelo, ou seja, pela variável preditora  $\hat{Y}$ . Ele é obtido através da equação

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

na qual  $i$  é o índice das observações na amostra ( $i = 1, \dots, n$ ),  $y_i$  são os valores observados da variável dependente,  $\hat{y}_i$  são os valores estimados e  $\bar{y}$  é a média dos  $y_i$  (Batista, J. L.

F., 2004). O  $R^2$  varia entre 0 e 1 e, de maneira geral, quanto maior seu valor melhor a concordância entre o modelo e os valores observados.

Como aponta Pala (2019), o  $R^2$  é uma métrica que pode apresentar problemas em pequenas amostras, não sendo recomendável seu cálculo em modelos com menos de cinquenta observações.

#### 4.2.2 Erro quadrático médio (MSE)

O erro quadrático médio (MSE, sigla para o termo em inglês *mean squared error*) é dado pela somatória do quadrado da diferença entre os valores observados  $y_i$  e os valores previstos  $\hat{y}_i$  em relação ao número total de elementos na amostra,  $n$ , como mostra a equação

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2.$$

O valor do MSE é interpretado como uma medida do desvio médio entre observado e predito. Quanto menor o valor do erro, melhor é a previsão do modelo.

#### 4.2.3 Erro absoluto médio (MAE)

O erro absoluto médio (MAE, sigla para o termo em inglês *mean absolut error*) fornece a verdadeira média dos desvios e possui a mesma unidade de medida da variável dependente. Ele é dado por

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|.$$



## 5 RESULTADOS DOS EXPERIMENTOS

Neste capítulo são apresentados os resultados dos experimentos realizados, tendo como objetivo a comparação do desempenho preditivo dos algoritmos de regressão selecionados.

Para a realização das simulações foi utilizado um computador Intel® Core™ i5-8265U CPU 1.60 GHz e 8,00 GB de memória RAM, utilizando o software Google Colab®, em linguagem de programação Python.

### 5.1 Desenho dos Experimentos

O conjunto de dados resultante da etapa de pré-processamento foi dividido em dois subconjuntos, um para treino e outro para teste, o primeiro com 80% do total de observações e o segundo com os 20% restantes.

Foram selecionados 5 algoritmos de AM além do modelo linear, conforme apresentado na Tabela 2, que mostra também os hiperparâmetros utilizados em cada modelo de AM (apenas aqueles diferentes do padrão de cada classe do *scikit-learn*<sup>1</sup>). Para definição dos hiperparâmetros dos modelos de AM, foi utilizada a ferramenta do *scikit-learn*, *Grid-SearchCV*, que realiza a busca pelos melhores parâmetros considerando a combinação de todos os parâmetros especificados (PEDREGOSA *et al.*, 2011).

Tabela 2 – Algoritmos de Aprendizado de Máquina e Hiperparâmetros selecionados.

Algoritmo	Hiperparâmetros
Árvore de Decisão (AD)	<i>random state=42, max depth=10, max features=1, min samples split= 6</i>
Floresta Aleatória (FA)	<i>random state = 42, max depth = 2, max features= 51, min samples leaf 26</i>
Máquinas de Vetores-Suporte (SVM)	<i>C = 7.0, epsilon = 0.02, kernel =“linear”</i>
Extra Tree (ET)	<i>random state=42, max depth=7, n estimators=50</i>
Gradient Boosting (GB)	<i>random state=42, max depth=1, min samples leaf= 41</i>

Fonte: Elaborada pela autora.

### 5.2 Resultados Gerais

De modo a verificar o efeito das variáveis escolhidas sobre o desmatamento, foram utilizados o modelo linear e as técnicas de aprendizado de máquina: árvores de decisão,

<sup>1</sup> <https://scikit-learn.org/stable/index.html>

florestas aleatórias, Extra Tree, Máquinas de Vetores-Suporte e Gradient Boosting. Para avaliar o desempenho preditivo dos modelos selecionados, foram utilizadas as medidas do coeficiente de determinação  $R^2$ , erro absoluto médio (MAE) e erro quadrático médio (MSE).

Para avaliar o impacto de cada variável independente na taxa de desmatamento, o primeiro modelo utilizado foi o modelo de regressão linear simples de mínimos quadrados ordinários que é tradicionalmente o mais utilizado na literatura relacionada. Neste método, procura-se minimizar a distância da resposta observada com os valores preditos. O modelo linear para todo o conjunto de dados utilizando o desmatamento como variável dependente resultou em um  $R^2$  igual a 0,5748. Tomando-se apenas o conjunto de treinamento, o valor do  $R^2$  é igual a 0,6397 e os resultados para as variáveis podem ser vistos na Tabela 3. A tabela está ordenada em valores crescentes de significância ( $P > |t|$ ).

A primeira coluna da Tabela 3, *coef*, é o coeficiente linear, o ponto em que a reta intercepta o eixo  $y$ . Para cada variável, é a medida de como uma mudança nessa variável afeta a variável independente. A coluna seguinte, *std err*, é o erro padrão estimado dos coeficientes de regressão, uma estimativa do desvio padrão do coeficiente. Na terceira coluna, o  $t$  é igual ao coeficiente dividido pelo erro padrão, é uma medida da precisão com que o coeficiente foi medido. A estatística  $t$  tem como hipótese nula que o coeficiente em questão é igual a zero. Um erro padrão baixo comparado a um coeficiente alto produz uma estatística  $t$  alta, o que significa uma significância alta para o seu coeficiente. A quarta coluna,  $P > |t|$  usa a estatística  $t$  para produzir o valor  $p$ , uma medida da probabilidade de que a variável não tenha efeito sobre a variável dependente, pois a hipótese nula é verdadeira. As últimas colunas, [0,025 e 0,975] representam os dois limites do coeficiente no intervalo de confiança de 95%. O intervalo de confiança fornece um conjunto de hipóteses plausíveis sobre o valor do parâmetro desconhecido. Se o valor da hipótese nula se encontrar neste intervalo, a hipótese não é rejeitada, caso contrário, ela será rejeitada.

Para Rivero *et al.* (2009), o método linear resultou que entre as 6 variáveis: efetivo bovino, produção de soja, arroz, milho, área plantada de culturas temporárias (excluindo arroz, milho e soja) e área plantada de culturas permanentes, a pecuária bovina, foi, individualmente, a principal causa imediata do desmatamento da Amazônia. Porém, considerando o nível de significância de 5%, pelos valores  $p$  calculados e apresentados na Tabela 3, as variáveis explicativas “Produção de soja”, “PIB Agropecuário”, “Despesa”, “Partido PSDB”, “Área plantada ou destinada”, “Produção de milho”, “Área colhida de Lavoura Permanente” e “Partido PHS” são estatisticamente significantes, indicando que essas variáveis têm efeito sobre a taxa de desmatamento.

Tabela 3 – Resultados da Regressão Linear para o conjunto de treinamento (Método dos mínimos quadrados ordinários)

Variável	coef	std err	t	P> t	[0,025	0,975]
Produção de soja	1475,08	357,20	4,130	0,000	768,02	2182,14
PIB Agropecuário	567,73	154,44	3,676	0,000	262,03	873,43
Despesa	-820,13	179,59	-4,567	0,000	-1175,62	-464,64
Partido PSDB	319,93	105,12	3,043	0,003	111,85	528,02
Área plantada ou destinada	-851,73	283,74	-3,002	0,003	-1413,38	-290,08
Produção milho	-1010,43	330,53	-3,057	0,003	-1664,69	-356,16
Área Colhida de Lavoura Perma-nente	478,70	210,32	2,276	0,025	62,39	895,01
Partido PHS	0,00	0,00	-2,088	0,039	0,00	0,00
Partido MDB	0,00	0,00	1,960	0,052	0,00	0,00
Valor da produção agropecuária	304,58	179,53	1,697	0,092	-50,78	659,95
Partido PFL	-163,72	102,55	-1,597	0,113	-366,71	39,26
Efetivo bovino	650,43	417,64	1,557	0,122	-176,26	1477,12
Partido PSB	-159,80	108,09	-1,478	0,142	-373,76	54,16
Partido PSD	194,52	132,65	1,466	0,145	-68,06	457,10
Partido PROS	186,47	127,61	1,461	0,147	-66,13	439,08
Fluxo de crédito ru-ral	-209,35	144,94	-1,444	0,151	-496,25	77,56
Partido PDT	-139,42	108,55	-1,284	0,201	-354,29	75,44
Partido PMDB	-111,20	120,10	-0,926	0,356	-348,92	126,52
Partido PMN	89,47	113,73	0,787	0,433	-135,65	314,58
Partido SD	-72,25	107,29	-0,673	0,502	-284,61	140,12
Partido PP	-49,15	105,29	-0,467	0,641	-257,56	159,25
Partido PPB	-48,74	105,37	-0,463	0,645	-257,31	159,84
Partido PSL	-42,84	105,42	-0,406	0,685	-251,52	165,84
Área Colhida de Lavoura Temporá-ria	151,62	380,53	0,398	0,691	-601,61	904,85
Partido PT	34,14	97,10	0,352	0,726	-158,07	226,35
Partido PCdoB	19,01	106,11	0,179	0,858	-191,02	229,04
Partido PSC	0,00	0,00	-0,153	0,878	0,00	0,00
Partido PPS	-17,34	117,79	-0,147	0,883	-250,51	215,82
Partido PTB	-14,57	104,99	-0,139	0,890	-222,39	193,25
Extensão de Ucs	10,16	203,25	0,050	0,960	-392,16	412,48
Focos de incêndio	-6,01	150,63	-0,040	0,968	-304,18	292,15
PIB	3,95	145,94	0,027	0,978	-284,93	292,82
População	-2,34	130,41	-0,018	0,986	-260,48	255,80

Fonte: Elaborada pela autora.

### 5.2.1 Modelos de aprendizado de máquina

A Tabela 4 apresenta um sumário dos resultados obtidos por cada modelo nas simulações realizadas com o conjunto de teste. O modelo que apresentou maior coeficiente de determinação e menor erro quadrático médio foi o FA, porém o menor erro absoluto médio foi obtido através do modelo ET. O modelo SVM também apresentou um bom desempenho geral.

Tabela 4 – Desempenho dos Modelos de aprendizado de máquina no conjunto de teste.

Modelo	$R^2$	MAE (km <sup>2</sup> )	MSE
Árvore de Decisão (AD)	0,5379	532,1	$5,775 \cdot 10^5$
Floresta Aleatória (FA)	0,6248	505,9	$4,689 \cdot 10^5$
SVM	0,6180	455,1	$4,774 \cdot 10^5$
Extra Tree (ET)	0,5845	442,6	$5,192 \cdot 10^5$
Gradient Boosting (GB)	0,5521	511,3	$5,597 \cdot 10^5$

Fonte: Elaborada pela autora.

O modelo de árvores de decisão e os comitês de árvores (FA, ET, GB) permitem quantificar a contribuição de cada variável preditiva. Os métodos baseados em árvore no *SciKit-Learn* possuem o recurso “importâncias das variáveis” (*feature importance*), um vetor de valores que relata a contribuição de cada variável para as previsões do modelo. As árvores de decisão individuais realizam intrinsecamente a seleção de variáveis, selecionando pontos de divisão apropriados. Quanto mais frequentemente uma variável é usada nos pontos de divisão de uma árvore, mais importante é essa variável. Nos conjuntos de árvores de decisão calcula-se a média das estimativas da capacidade preditiva de várias árvores aleatórias para definir a importância (PEDREGOSA *et al.*, 2011). É importante ressaltar que as importâncias dos atributos são invariavelmente não negativas, consequentemente, elas não transmitem informações sobre a correlação positiva ou negativa entre um preditor e a variável alvo.

Na Figura 21, são apresentadas as importâncias para os modelos baseados em árvore. Apenas as cinco variáveis mais relevantes de cada modelo são mostradas, ordenadas por importância. Para o modelo ET, a variável “Área colhida de lavoura temporária” foi o atributo de maior importância, representando 19% do poder preditivo do algoritmo, como mostra a Figura 21b. Para os modelos AD, FA e GB, é a variável “Área colhida de lavoura permanente” a mais relevante. Para o modelo AD, Figura 21a, a importância é de 47%, para o modelo FA, Figura 21c, ela possui relevância de 86% e para o modelo GB, Figura 21d, a relevância é de 52%

O modelo FA foi o que resultou em menos variáveis com importância diferente de zero, apenas as cinco apresentados na Figura 21c.



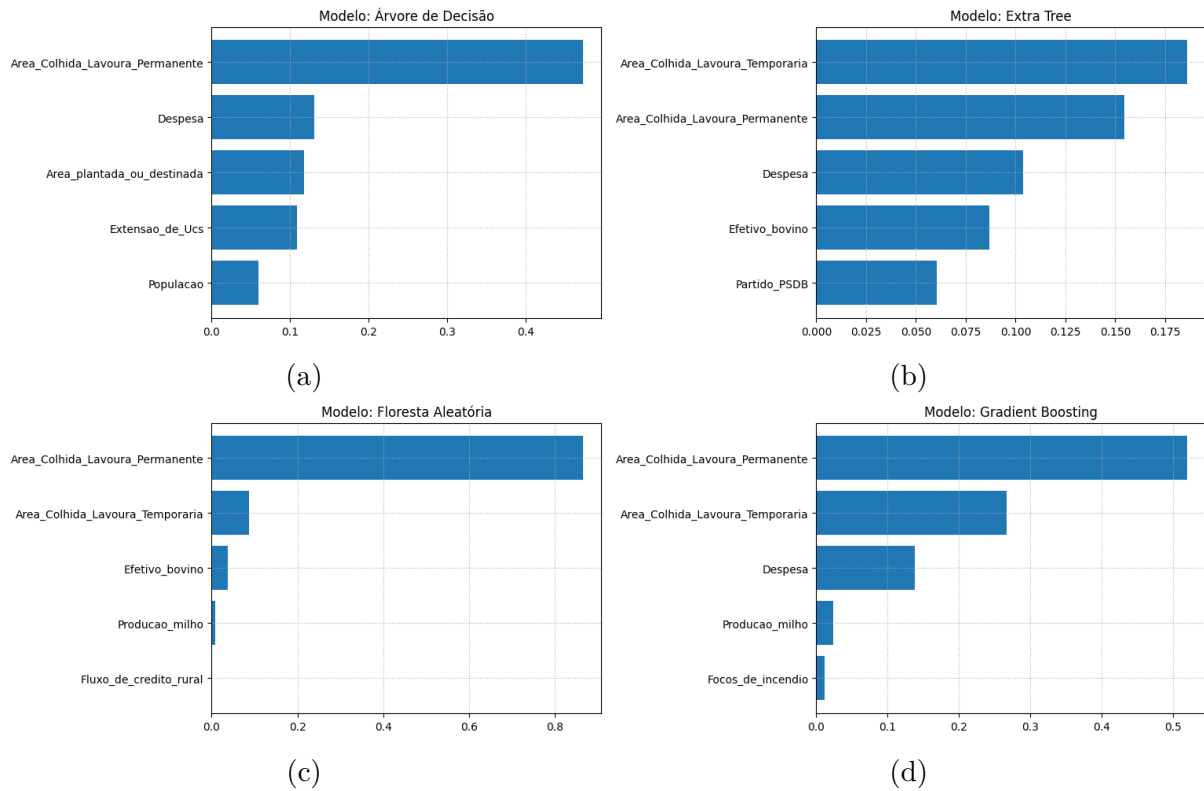


Figura 21 – Importância das variáveis para os modelos de árvore.

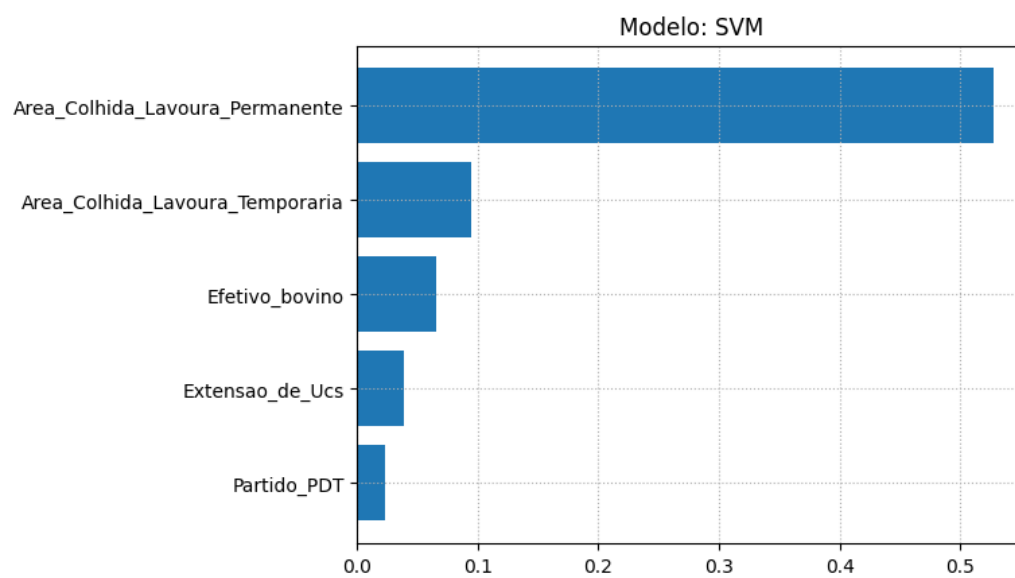
A variável “Produção de soja” teve importância diferente de zero para os modelos ET e GB, com 5% e 1%, respectivamente. A variável “Efetivo Bovino” teve importância diferente de zero para os modelos AD, FA e ET, com importâncias relativas de 1%, 4% e 9%, respectivamente. Já a variável “Produção de milho” teve importância diferente de zero para os modelos FA, ET e GB com 1%, 5% e 2% respectivamente.

Para o modelo SVM existe o recurso do *SciKit-Learn* que permite obter a importância de cada variável utilizando o método de permutação (*Permutation feature importance*). Esse método mede o aumento no erro do modelo se os valores da variável forem permutados. A Figura 22 mostra as cinco variáveis com maior importância de permutação para este modelo.

Assim como para os modelos AD, FA e GB, a variável “Área Colhida de Lavoura Permanente” apresentou maior importância (57%). A variável “Efetivo Bovino” teve importância de 6%, “Produção de milho” 0,4% e a “Produção de soja” 0,2% apenas.

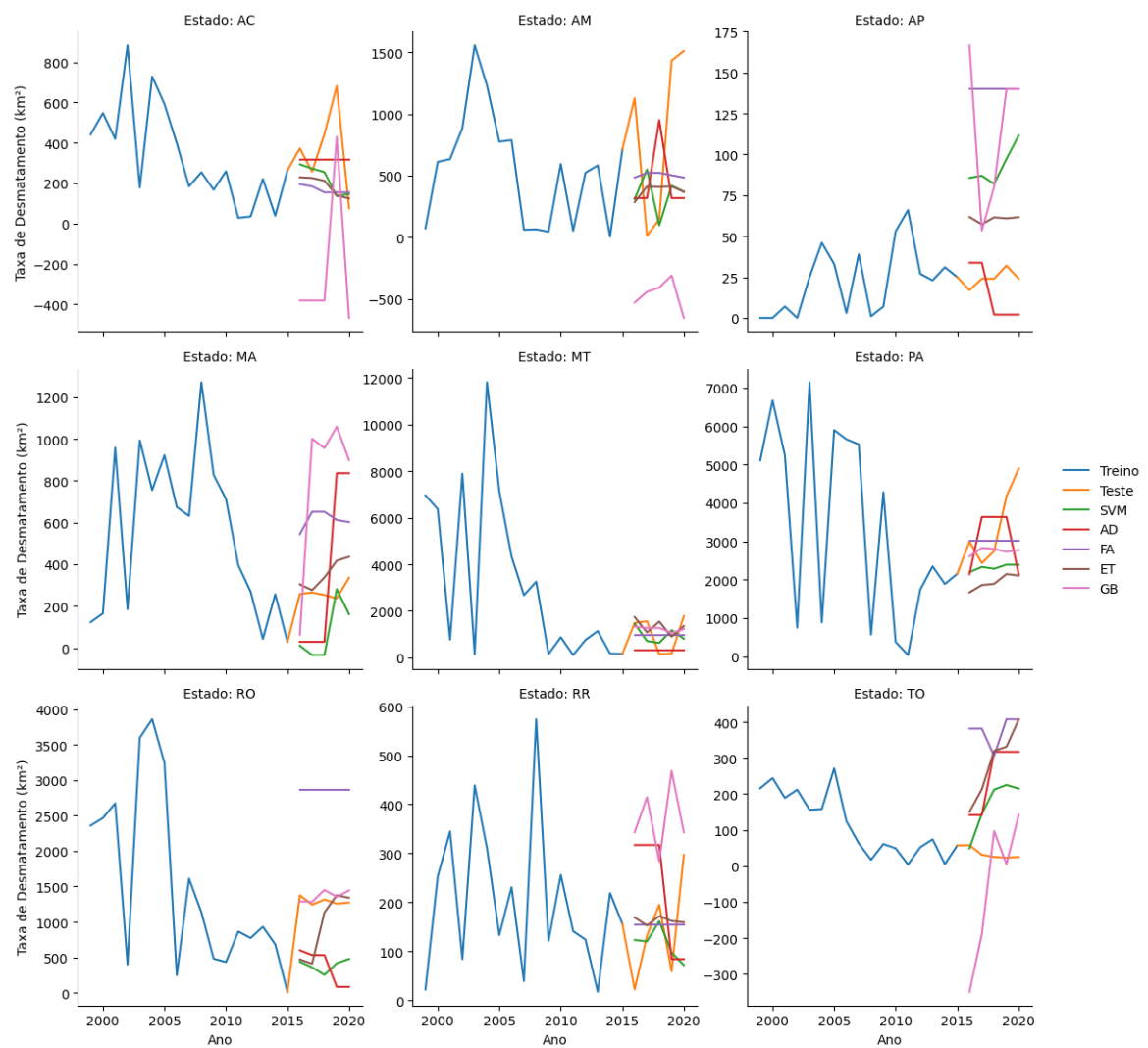
A Figura 23 mostra a variável “Taxa de Desmatamento” para todos os estados estudados. Na figura é possível distinguir os dados usados para treino e teste, assim como os valores previstos pelos cinco algoritmos de AM. O algoritmo GB resultou em muitos valores negativos para esta variável, especialmente para o Amazonas, cujos valores previstos foram todos negativos.

Figura 22 – Importância das variáveis para o modelo SVM.



Fonte: Elaborado pela autora.

Figura 23 – Taxa de desmatamento para cada estado: treino, teste e valores previstos pelos modelos de AM.



Fonte: Elaborado pela autora.



## 6 CONCLUSÃO

Neste trabalho, o desempenho preditivo de cinco algoritmos de aprendizado de máquina foi avaliado no estudo da influência de diferentes variáveis socioeconômicas para o desmatamento nos sete estados que abrangem a Amazônia brasileira, mais os estados do Mato Grosso e Maranhão, cuja presença do bioma Amazônia é parcial, porém foram considerados em sua totalidade. Os dados em painel foram reunidos a partir de diferentes fontes (IBGE, Ipeadata, Instituto Sócio Ambiental e Wikipédia) para o período 1999-2020, resultando em um conjunto de dados em painel com 198 linhas e 16 variáveis.

O modelo de regressão linear simples pelo método dos mínimos quadrados ordinários foi utilizado para analisar a significância e influência das variáveis escolhidas para o desmatamento. Este modelo é tradicionalmente utilizado na literatura para analisar os conjuntos de dados selecionados. Diferentemente dos resultados de Rivero *et al.* (2009), que utilizaram apenas alguns municípios e um período menor e concluíram que o número de cabeças de gado é a principal causa de desmatamento na Amazônia, no presente trabalho, a variável que apresentou maior significância no modelo e poderia ser indicada como uma das principais causas diretas no desmatamento foi a produção de soja. No entanto, esta diferença pode ser atribuída ao uso da área total do estado do Mato Grosso, onde a produção de soja apresenta os maiores números. Sabe-se que no Mato Grosso encontra-se três tipos de biomas (Amazônia, Cerrado e Pantanal), porém neste trabalho não foi feita esta distinção. O efetivo bovino, apesar de significância superior à 0,005 tem coeficiente de 650, o que representa o segundo maior impacto nominal na variável desmatamento entre todos os atributos, atrás apenas da produção de soja, que apresentou coeficiente de 1475.

Entre os partidos políticos com governadores em exercício no período analisado, apenas o PSDB teve significância menor do que 0,005 e coeficiente positivo de 319. Os demais partidos tiveram valores de significância igual ou superior à 0,04.

Entre os algoritmos de AM, foram utilizados quatro modelos de árvore: Árvore de Decisão, Floresta Aleatória, Extra Tree e Gradient Boosting, além do modelo Máquinas de Vetores-Suporte (SVM). Entre os modelos de árvore, teve melhor desempenho geral o Floresta Aleatória, com maior coeficiente de determinação  $R^2$  e menor erro quadrático médio. O modelo SVM foi o segundo melhor modelo em termos das medidas de desempenho. Foi utilizado o algoritmo de otimização de hiper parâmetros *GridSearchCV*, do *Scikit Learn* (PEDREGOSA *et al.*, 2011), para todos os modelos. Os algoritmos de aprendizado de máquina baseados em árvore possuem a vantagem de serem mais fáceis de interpretar em comparação com outros modelos e são capazes de lidar com uma grande quantidade de características, identificando as mais relevantes para a previsão.

A verificação das importâncias das variáveis na previsão dos modelos de Aprendizado de Máquina, teve, principalmente, a área colhida de lavoura permanente como variável mais relevante para o desmatamento, exceto para o algoritmo Extra Tree, que no entanto, teve essa variável como a segunda mais importante. São permanentes as culturas que não demandam novo plantio, permitindo colheitas sucessivas como, por exemplo, café, maçã, pera, uva, manga, laranja etc (IBGE Educa, 2023). A área colhida de lavoura temporária, culturas de curta ou média duração, que após a colheita necessitam de novo plantio para produzir, como, por exemplo: soja, milho, feijão etc, aparece como a segunda mais importante nos modelos SVM, Floresta aleatória e Gradient Boosting.

Vale destacar a presença da área de lavoura permanente nas importâncias associadas aos modelos de aprendizado de máquina, uma vez que, de acordo com os dados do IBGE, esse tipo de lavoura apresentou redução ao longo dos últimos anos na maior parte dos estados estudados e a área de lavoura temporária apresentou crescimento. De acordo com Rivero *et al.* (2009), a soja, principal cultura temporária em crescimento no país, é cultivada em áreas anteriormente ocupadas por outros tipos de culturas ou pela pecuária. Ou seja, ao longo dos últimos anos houve substituição das áreas destinadas à culturas permanentes e pecuária pela soja. Isso explica o menor impacto das lavouras temporárias e, conseqüentemente, da soja, no desmatamento, uma vez que essas lavouras ocupam áreas já desmatadas.

A variável focos de incêndio também não se mostrou relevante para os modelos testados, o que corrobora os resultados de Jakimow *et al.* (2023), que mostram que maior parte dos focos de incêndio registrados ocorreram em áreas já desmatadas em relação ao período considerado.

Entre os modelos estudados, o modelo Gradient Boosting, apesar não ter tido um desempenho ruim em relação aos demais modelos no que diz respeito às medidas de coeficiente de determinação, erro quadrático médio e erro absoluto médio, resultou em muitos valores negativos para a variável desmatamento nas suas previsões.

Para trabalhos futuros, o conjunto de dados utilizado pode ser aprimorado, reunindo-se dados referentes estritamente ao Bioma Amazônia, por exemplo, e fazendo-se uma seleção quanto aos atributos efetivamente relevantes para o problema. Neste trabalho, verificou-se, por exemplo, que variáveis como PIB e População não apresentaram relevância para a variável desmatamento no período estudado.

É possível ainda, ampliar o período da análise, incluindo períodos de intenso crescimento nas atividades agropecuárias na região.

Futuramente também, pode-se implementar outros algoritmos de Aprendizado de Máquina e avaliar seus desempenhos nesta análise, como redes neurais e XGBoost, por exemplo.

## REFERÊNCIAS

Agência IBGE Notícias. **PAM 2011: valor da produção agrícola cresce 27,1**<https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/14303-asi-pam-2011-valor-da-producao-agricola-cresce-271-em-relacao-a-2010>. Acesso em: 12 set. 2023.

Agência IBGE Notícias. **Valor de produção bate recorde, mas safra 2021 não supera ano anterior**. 2022. Disponível em: <https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/34943-valor-de-producao-bate-recorde-mas-safra-2021-nao-supera-ano-anterior>. Acesso em: 12 set. 2023.

Banco Central do Brasil. **Matriz de Dados do Crédito Rural - Crédito Concedido**. 2023. Disponível em: <https://www.bcb.gov.br/estabilidadefinanceira/micrrural>. Acesso em: 13 set. 2023.

Batista, J. L. F. **Análise de Regressão Aplicada**. 2004. Disponível em: <http://cmq.esalq.usp.br/wiki/lib/exe/fetch.php?media=biometria:encontros:applied-regression.pdf>. Acesso em: 07 nov. 2023.

BISHOP, C. M. **Pattern Recognition and Machine Learning**. Springer, 2006. v. 4. 738 p. (Information science and statistics, 4). ISSN 10179909. ISBN 9780387310732. Disponível em: <http://www.library.wisc.edu/selectedtocs/bg0137.pdf>.

CHEN, W. *et al.* Mapping a burned forest area from landsat tm data by multiple methods. **Geomatics, Natural Hazards and Risk**, v. 7, n. 1, p. 384–402, 2016. Disponível em: <https://doi.org/10.1080/19475705.2014.925982>.

CHUVIECO, E. *et al.* Historical background and current developments for mapping burned area from satellite earth observation. **Remote Sensing of Environment**, v. 225, p. 45–64, 2019. ISSN 0034-4257. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0034425719300689>.

CNA. **PIB do Agronegócio alcança participação de 26,6% no PIB brasileiro em 2020**. 2023. Disponível em: <https://cnabrazil.org.br/publicacoes/pib-do-agronegocio-alcanca-participacao-de-26-6-no-pib-brasileiro-em-2020>. Acesso em: 13 set. 2023.

COHEN, W. B.; GOWARD, S. N. Landsat's Role in Ecological Applications of Remote Sensing. **BioScience**, v. 54, n. 6, p. 535–545, 06 2004. ISSN 0006-3568. Disponível em: [https://doi.org/10.1641/0006-3568\(2004\)054\[0535:LRIEAO\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2004)054[0535:LRIEAO]2.0.CO;2).

CÂMARA, G. *et al.* Impact of land tenure on deforestation control and forest restoration in brazilian amazonia. **Environ. Res. Lett.**, v. 18, p. 065005, 2023.

das Neves, P. B. T. *et al.* Amazon rainforest deforestation influenced by clandestine and regular roadway network. **Land Use Policy**, v. 108, p. 105510, 2021. ISSN 0264-8377. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0264837721002337>.

ELACHI, C.; ZYL, J. van. Appendix a: Use of multiple sensors for surface observations. *In: \_\_\_\_\_*. **Introduction to the Physics and Techniques of Remote Sensing**. John Wiley & Sons, Ltd, 2021. p. 507–510. ISBN 9781119523048. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119523048.app1>.

HITCHCOCK, H.; HOFFER, R. Mapping a recent forest fire with erts-1 mss data. **Conf on Earth Resour Obs and Inf Anal Syst, Remote Sensing of Earth Resour**, v. 3, p. 449 – 461, 1974. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0016217081&partnerID=40&md5=4829b6798ee6a1cd2e31dabb087165aa>.

IBF. **Bioma Amazônico**. 2020. Disponível em: <https://www.ibflorestas.org.br/bioma-amazonico>. Acesso em: 10 mar. 2023.

IBGE. **IBGE Explica**. 2023. Disponível em: <https://www.ibge.gov.br/explica/pib.php>. Acesso em: 11 set. 2023.

IBGE. **Sensoriamento remoto**. 2023. Disponível em: <https://atlasescolar.ibge.gov.br/conceitos-gerais/o-que-e-cartografia/sensoriamento-remoto.html>. Acesso em: 22 fev. 2023.

IBGE Educa. **Culturas temporárias e Culturas permanentes**. 2023. Disponível em: <https://educa.ibge.gov.br/professores/educa-atividades/17666-culturas-temporarias-e-culturas-permanentes.html>. Acesso em: 07 nov. 2023.

INPE. **BDQUEIMADAS**. Disponível em: <https://queimadas.dgi.inpe.br/>. Acesso em: 22 fev. 2023.

INPE. **PRODES**. Disponível em: <http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes>. Acesso em: 11 jul. 2023.

INPE. **Programa Queimadas**. Disponível em: <https://queimadas.dgi.inpe.br/queimadas/portal/informacoes/perguntas-frequentes>. Acesso em: 22 fev. 2023.

Instituto SocioAmbiental. **Sobre**. 2023. Disponível em: <https://uc.socioambiental.org/pt-br/sobre>. Acesso em: 13 set. 2023.

IPEA. **Ipeadata**. 2023. Disponível em: <http://ipeadata.gov.br/Default.aspx>. Acesso em: 17 de jun. 2023.

JAKIMOW, B. *et al.* Deforestation and agricultural fires in south-west pará, brazil, under political changes from 2014 to 2020. **Journal of Land Use Science**, Taylor and Francis, v. 18, n. 1, p. 176–195, 2023. Disponível em: <https://doi.org/10.1080/1747423X.2023.2195420>.

KIM, S. J. *et al.* Multi-temporal analysis of forest fire probability using socio-economic and environmental variables. **Remote Sensing**, v. 11, n. 1, 2019. ISSN 2072-4292. Disponível em: <https://www.mdpi.com/2072-4292/11/1/86>.

LEUENBERGER, M. *et al.* Wildfire susceptibility mapping: Deterministic vs. stochastic approaches. **Environmental Modelling & Software**, v. 101, p. 194–203, 2018. ISSN 1364-8152. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1364815217303316>.



- MITCHELL, T. M. **Machine learning**. [S.l.: s.n.]: McGraw-hill New York, 1997. v. 1.
- Murer, B. M and Futada, S. M. **Painel de Dados**. 2023. Disponível em: <https://uc.socioambiental.org/pt-br/paineldedadoshistorico>. Acesso em: 13 set. 2023.
- NASA. **Landsat Science**. 2023. Disponível em: <https://landsat.gsfc.nasa.gov/about/landsat-timeline/>. Acesso em: 22 fev. 2023.
- OLIVEIRA, S. *et al.* Modeling spatial patterns of fire occurrence in mediterranean europe using multiple regression and random forest. **Forest Ecology and Management**, v. 275, p. 117–129, 2012. ISSN 0378-1127. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0378112712001272>.
- PALA, L. O. O. **Revisitando a estimação do Coeficiente de Determinação**. 2019. 113 p. Dissertação (Mestrado em Estatística Aplicada e Biometria) — Programa de Pós-Graduação em Estatística Aplicada e Biometria, Universidade Federal de Alfenas, Minas Gerais, 2019.
- PEDREGOSA, F. *et al.* Scikit-learn: Machine learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011.
- RIVERO, S. *et al.* Pecuária e desmatamento: uma análise das principais causas diretas do desmatamento na amazônia. **Nova Economia**, Nova Economia, v. 19, n. 1, p. 41–66, Jan 2009. ISSN 0103-6351. Disponível em: <https://doi.org/10.1590/S0103-63512009000100003>.
- ROMERO, R. **Comitês de Árvores de Decisão -Videoaula 09, Professora Roseli Romero**. São Carlos: ICMC, USP, 2022. Disponível em: <https://cursosextensao.usp.br/mod/page/view.php?id=178418&forceview=1>. Acesso em: 22 mai. 2023.
- SHEPPARD, K. **Data Formats for Panel Data Analysis**. 2021. <https://bashtage.github.io/linearmodels/panel/examples/data-formats.html>. Acesso em: 19 set. 2023.
- THACH, N. N. *et al.* Spatial pattern assessment of tropical forest fire danger at thuan chau area (vietnam) using gis-based advanced machine learning algorithms: A comparative study. **Ecological Informatics**, v. 46, p. 74–85, 2018. ISSN 1574-9541. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1574954118300852>.
- VERDÉRIO, A. **Sobre o uso de regressão por vetores suporte para a construção de modelos em um método de região de confiança sem derivadas**. 2015. 131 p. Tese (Doutorado em Matemática) — Universidade Federal do Paraná, Setor de Ciências Exatas, Curitiba, 2015. Disponível em: <https://acervodigital.ufpr.br/xmlui/bitstream/handle/1884/37833/R%20-%20T%20-%20ADRIANO%20VERDERIO.pdf>. Acesso em: 04 nov. 2023.
- VILAR, L. *et al.* Multitemporal modelling of socio-economic wildfire drivers in central spain between the 1980s and the 2000s: Comparing generalized linear models to machine learning algorithms. **PLOS ONE**, Public Library of Science, v. 11, n. 8, p. 1–17, 08 2016. Disponível em: <https://doi.org/10.1371/journal.pone.0161344>.
- VOLPATO, M. *et al.* Fire foci and their spatiotemporal relations to weather variables and land uses in the state of mato grosso. **Environ. Dev. Sustain.**, p. 1573–2975, 2022. Disponível em: <https://doi.org/10.1007/s10668-022-02573-3>.

WANG, Y.; HUANG, P. Potential fire risks in south america under anthropogenic forcing hidden by the atlantic multidecadal oscillation. **Nat Commun**, v. 13, p. 2437, 2022.

WIKIPÉDIA. **Lista de governadores de Rondônia — Wikipédia, a enciclopédia livre**. 2023. [Acesso em: 15 set. 2023]. Disponível em: [https://pt.wikipedia.org/wiki/Lista\\_de\\_governadores\\_de\\_Rond%C3%B4nia](https://pt.wikipedia.org/wiki/Lista_de_governadores_de_Rond%C3%B4nia).

WIKIPÉDIA. **Lista de governadores de Roraima — Wikipédia, a enciclopédia livre**. 2023. [Acesso em: 15 set. 2023]. Disponível em: [https://pt.wikipedia.org/wiki/Lista\\_de\\_governadores\\_de\\_Roraima](https://pt.wikipedia.org/wiki/Lista_de_governadores_de_Roraima).

WIKIPÉDIA. **Lista de governadores do Acre — Wikipédia, a enciclopédia livre**. 2023. [Acesso em: 15 set. 2023]. Disponível em: [https://pt.wikipedia.org/wiki/Lista\\_de\\_governadores\\_do\\_Acre](https://pt.wikipedia.org/wiki/Lista_de_governadores_do_Acre).

WIKIPÉDIA. **Lista de governadores do Amapá — Wikipédia, a enciclopédia livre**. 2023. [Acesso em: 15 set. 2023]. Disponível em: [https://pt.wikipedia.org/w/index.php?title=Lista\\_de\\_governadores\\_do\\_Amap%C3%A1&oldid=65190123](https://pt.wikipedia.org/w/index.php?title=Lista_de_governadores_do_Amap%C3%A1&oldid=65190123).

WIKIPÉDIA. **Lista de governadores do Amazonas — Wikipédia, a enciclopédia livre**. 2023. [Acesso em: 15 set. 2023]. Disponível em: [https://pt.wikipedia.org/wiki/Lista\\_de\\_governadores\\_do\\_Amazonas](https://pt.wikipedia.org/wiki/Lista_de_governadores_do_Amazonas).

WIKIPÉDIA. **Lista de governadores do Maranhão — Wikipédia, a enciclopédia livre**. 2023. [Acesso em: 15 set. 2023]. Disponível em: [https://pt.wikipedia.org/wiki/Lista\\_de\\_governadores\\_do\\_Maranh%C3%A3o](https://pt.wikipedia.org/wiki/Lista_de_governadores_do_Maranh%C3%A3o).

WIKIPÉDIA. **Lista de governadores do Mato Grosso — Wikipédia, a enciclopédia livre**. 2023. [Acesso em: 15 set. 2023]. Disponível em: [https://pt.wikipedia.org/wiki/Lista\\_de\\_governadores\\_de\\_Mato\\_Grosso](https://pt.wikipedia.org/wiki/Lista_de_governadores_de_Mato_Grosso).

WIKIPÉDIA. **Lista de governadores do Pará — Wikipédia, a enciclopédia livre**. 2023. [Acesso em: 15 set. 2023]. Disponível em: [https://pt.wikipedia.org/wiki/Lista\\_de\\_governadores\\_do\\_Par%C3%A1](https://pt.wikipedia.org/wiki/Lista_de_governadores_do_Par%C3%A1).

WIKIPÉDIA. **Lista de governadores do Tocantins — Wikipédia, a enciclopédia livre**. 2023. [Acesso em: 15 set. 2023]. Disponível em: [https://pt.wikipedia.org/wiki/Lista\\_de\\_governadores\\_do\\_Tocantins](https://pt.wikipedia.org/wiki/Lista_de_governadores_do_Tocantins).

WOOSTER, M. J. *et al.* Satellite remote sensing of active fires: History and current status, applications and future requirements. **Remote Sensing of Environment**, v. 267, p. 112694, 2021. ISSN 0034-4257. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0034425721004144>.

YOUNG, C. E. F. Public policies and deforestation in the brazilian amazon. **Planejamento e Políticas Públicas**, n. 18, ago. 2022. Disponível em: <https://www.ipea.gov.br/ppp/index.php/PPP/article/view/103>.