

# **Evolução do Desmatamento no Brasil**

## **Integrantes:**

- Lucas Magliari da Purificação Pereira
- Matheus Esteves Pereira Affonso
- Leonardo Bastos Yuan Gouvea
- Dione Lucas Souza da Costa



Universidade Presbiteriana  
**Mackenzie**

## **Introdução:**

O desmatamento no Brasil, em especial na Amazônia, é um dos maiores desafios ambientais e socioeconômicos globais. Entre 1988 e 2022, mais de 800 mil km<sup>2</sup> de floresta amazônica foram desmatados, impactando não apenas a biodiversidade local, mas também o equilíbrio climático global. A Amazônia, que concentra 60% das florestas tropicais do mundo, desempenha um papel vital na regulação do clima, servindo como um dos maiores sumidouros de carbono do planeta. Dados do INPE revelam um crescimento de 56,6% no desmatamento entre 2019 e 2021, uma tendência preocupante que aponta para a necessidade de medidas urgentes. Esta pesquisa visa contextualizar e analisar a evolução do desmatamento na região, abordando as principais causas e consequências desse processo, bem como a criação de soluções preditivas que ajudem a mitigar os seus impactos no futuro. A área de conhecimento envolvida inclui ciências ambientais, modelagem estatística e machine learning aplicados à análise geoespacial.

## **Motivações e Justificativa:**

A relevância deste estudo se justifica pela magnitude dos danos causados pelo desmatamento, que afeta o clima, a biodiversidade e a vida de milhões de pessoas que dependem direta ou indiretamente da floresta. O aumento dos índices de desmatamento não apenas agrava o aquecimento global, mas também afeta a economia local, principalmente no que diz respeito à produção de água, agricultura sustentável e turismo. Estudos indicam que, se mantido o ritmo atual, a Amazônia poderá atingir um ponto de não retorno, tornando-se uma savana degradada. A proposta de criar um modelo preditivo para os índices de desmatamento visa fornecer ferramentas para antecipar tendências futuras, permitindo que políticas públicas e estratégias de conservação sejam implementadas de maneira mais eficiente e proativa. Além disso, o modelo pode servir como base para estudos de viabilidade de programas de compensação de carbono e monitoramento ambiental em tempo real.

## **Objetivo:**

**Objetivo Geral:** Analisar a evolução do desmatamento no Brasil e desenvolver um modelo preditivo capaz de antecipar futuros índices de desmatamento, com foco na região amazônica, contribuindo para a criação de políticas de mitigação.

## **Objetivos Específicos:**

1. Analisar a evolução temporal e espacial do desmatamento no Brasil entre 1988 e 2022, com ênfase na Amazônia Legal.

2. Identificar os fatores socioeconômicos e ambientais que influenciam o desmatamento, como atividades agropecuárias, mineração e mudanças nas políticas de preservação.
3. Desenvolver e validar um modelo preditivo utilizando técnicas de machine learning, com base em dados históricos e variáveis externas, capaz de prever os índices de desmatamento em diferentes regiões.
4. Analisar os impactos ambientais do desmatamento, como perda de biodiversidade e emissões de carbono, e utilizar os resultados do modelo preditivo para sugerir intervenções estratégicas.
5. Propor recomendações para políticas públicas de combate ao desmatamento, que integrem os resultados das previsões e os padrões históricos.

### **Proposta de Modelo Preditivo:**

Para prever os índices de desmatamento, propomos o desenvolvimento de um modelo de machine learning utilizando algoritmos como Random Forest, XGBoost ou redes neurais artificiais, que são eficazes em prever tendências com base em dados complexos e de grande escala. O modelo será treinado e validado com dados históricos de desmatamento fornecidos pelo INPE, incluindo variáveis como:

- **Área desmatada por estado e ano:** Indicando a extensão de floresta perdida ao longo dos anos em cada região.
- **Atividades econômicas:** Dados sobre agricultura, pecuária, mineração e extração de madeira, que são os principais motores do desmatamento.
- **Variáveis ambientais:** Incluindo precipitação, temperatura e umidade, que podem influenciar a vulnerabilidade de certas áreas ao desmatamento.
- **Políticas públicas:** Mudanças em legislações ambientais, presença de fiscalizações e zonas de proteção.
- **Fatores socioeconômicos:** Níveis de pobreza, crescimento populacional, expansão urbana e incentivos econômicos.

Esses dados serão processados para criar um modelo capaz de prever as áreas mais propensas ao desmatamento em curto, médio e longo prazo. O modelo usará uma abordagem de séries temporais, combinada com análise geoespacial, permitindo identificar hotspots de desmatamento com base em padrões históricos e mudanças nos fatores socioeconômicos.

### **Etapas do Desenvolvimento do Modelo:**

1. **Coleta e Limpeza de Dados:** Organizar e limpar os dados históricos, garantindo a uniformidade e preenchimento adequado de lacunas.
2. **Exploração de Dados:** Análise exploratória para identificar padrões e correlações entre as variáveis.
3. **Escolha do Algoritmo:** Testar diferentes algoritmos (Random Forest, XGBoost, redes neurais) e escolher o que melhor se adapta aos dados.
4. **Treinamento do Modelo:** Treinar o modelo com um conjunto de dados de treino (70% dos dados) e validar com um conjunto de teste (30% dos dados).
5. **Validação e Ajustes:** Validar a precisão do modelo e ajustar hiperparâmetros para melhorar o desempenho.
6. **Interpretação dos Resultados:** Gerar previsões e mapear as áreas de maior risco de desmatamento para os próximos anos.
7. **Implementação em Sistemas de Monitoramento:** Integrar o modelo em plataformas de monitoramento ambiental para oferecer previsões em tempo real.

O modelo preditivo será capaz de fornecer previsões em nível estadual e, eventualmente, até em nível municipal, com margens de erro adequadas para o planejamento de intervenções rápidas e eficazes. Além disso, ele poderá ser ajustado e melhorado continuamente à medida que novos dados forem coletados, permitindo um monitoramento mais dinâmico e preciso.

#### **Descrição da Base de Dados:**

A base de dados utilizada no estudo é composta por informações sobre o desmatamento na Amazônia Legal, abrangendo os estados de Acre, Amazonas, Amapá, Maranhão, Mato Grosso, Pará, Rondônia, Roraima e Tocantins. A coleta de dados foi realizada pelo INPE através de monitoramento por satélite, utilizando tecnologia de sensoriamento remoto para medir com precisão a extensão do desmatamento ao longo dos anos. Os principais atributos são:

- **Ano de Referência:** Cada registro identifica o ano em que a medição foi realizada, permitindo a análise temporal do desmatamento.
- **Estados:** A base é dividida por estado, facilitando a análise espacial e permitindo a identificação de padrões regionais de desmatamento.
- **Área Total Desmatada:** Medida em quilômetros quadrados (km<sup>2</sup>), representando a quantidade total de floresta derrubada.

- **Fatores Econômicos:** Inclui dados sobre atividades agrícolas, mineração, pecuária e outros setores econômicos que influenciam o desmatamento.
- **Variáveis Climáticas:** Informações sobre precipitação, temperatura média e estações secas, que podem aumentar a vulnerabilidade de áreas ao desmatamento.

Os dados históricos a seguir são coletados a partir do PRODES, um projeto da INPE (Instituto Nacional de Pesquisas

Espaciais) - <http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes>

## Referencial Teórico

O desmatamento na Amazônia é um tema amplamente estudado devido aos seus impactos ambientais globais. Silva (2020) investigou os efeitos ambientais do desmatamento no Brasil, destacando a perda de biodiversidade e o aumento das emissões de gases de efeito estufa. O estudo enfatiza a necessidade de monitoramento contínuo e de políticas públicas eficazes para a conservação florestal.

No contexto de modelagem preditiva, Ferreira (2021) desenvolveu um modelo estatístico utilizando regressão linear para prever as taxas de desmatamento com base em variáveis socioeconômicas. Embora o modelo tenha fornecido insights valiosos, apresentou limitações na captura de relações não lineares complexas entre as variáveis. Para superar essas limitações, técnicas de aprendizado de máquina como Random Forest e XGBoost têm sido exploradas, pois conseguem lidar com grandes conjuntos de dados e capturar interações complexas.

Redes Neurais Artificiais também foram aplicadas em estudos que utilizaram essa técnica para prever áreas de risco de desmatamento. Apesar da alta capacidade de modelagem, essas redes demandam grande volume de dados e alto poder computacional, o que pode ser uma limitação em alguns cenários.

As vantagens das técnicas de aprendizado de máquina incluem maior precisão preditiva e a capacidade de lidar com dados heterogêneos. No entanto, enfrentam desafios como a necessidade de dados de alta qualidade e

interpretabilidade dos modelos. A seleção adequada de variáveis e o préprocessamento eficiente dos dados são essenciais para o sucesso desses modelos.

Os principais conceitos envolvidos na solução proposta são:

- **Desmatamento:** Remoção da cobertura vegetal natural, resultando em perda de biodiversidade e alterações climáticas.
- **Aprendizado de Máquina:** Conjunto de técnicas que permitem que os computadores aprendam padrões a partir de dados, sem serem explicitamente programados para tal.
- **Modelagem Preditiva:** Uso de modelos estatísticos e algoritmos para prever resultados futuros com base em dados históricos.
- **Análise Geoespacial:** Estudo de fenômenos com referência espacial, fundamental para mapear e entender a distribuição do desmatamento.

A proposta deste trabalho é desenvolver um modelo preditivo robusto que utilize técnicas avançadas de aprendizado de máquina para prever as tendências de desmatamento, superando as limitações de estudos anteriores e contribuindo para ações de mitigação mais eficazes.

## Pipeline da Solução

A solução proposta seguirá um pipeline estruturado em etapas sequenciais:

### 1. Coleta e Limpeza de Dados:

- Reunir dados históricos de desmatamento fornecidos pelo INPE e outras fontes relevantes.
- Incluir variáveis como área desmatada, atividades econômicas, dados climáticos e políticas públicas.
- Realizar limpeza dos dados para remover inconsistências e tratar valores ausentes.

### 2. Análise Exploratória de Dados:

- Analisar estatisticamente os dados para identificar padrões, tendências e outliers.
- Visualizar a distribuição espacial e temporal do desmatamento.
- Identificar correlações entre as variáveis independentes e a variável dependente (desmatamento).

### 3. Pré-processamento dos Dados:

- Normalizar ou padronizar variáveis numéricas.
- Codificar variáveis categóricas utilizando técnicas como one-hot encoding.
- Dividir o conjunto de dados em treino e teste (e possivelmente validação).

#### 4. Seleção e Treinamento do Modelo:

- Testar algoritmos como Random Forest, XGBoost e redes neurais artificiais.
- Utilizar validação cruzada para avaliar o desempenho dos modelos.
  - Selecionar o modelo com melhor desempenho com base em métricas como RMSE, MAE e  $R^2$ .

#### 5. Validação e Ajuste do Modelo:

- Otimizar hiperparâmetros utilizando técnicas como Grid Search ou Random Search.
- Avaliar o modelo em dados de teste para verificar a capacidade de generalização.

#### 6. Interpretação e Visualização dos Resultados:

- Gerar mapas e gráficos que demonstrem as previsões de desmatamento.
- Identificar áreas de alto risco e possíveis fatores causais.

#### 7. Implementação da Solução:

- Integrar o modelo em uma plataforma ou sistema de monitoramento.
- Proporcionar atualizações periódicas com novos dados para manter o modelo relevante.

#### 8. Revisão e Atualização Contínua:

- Monitorar o desempenho do modelo ao longo do tempo.
- Atualizar o modelo com novos dados e refinar conforme necessário.

Este pipeline permite uma abordagem sistemática para desenvolver um modelo preditivo eficaz, garantindo que todas as etapas, desde a coleta de dados até a implementação, sejam cuidadosamente executadas.

#### Cronograma

Data de Entrega	Atividade
-----------------	-----------

30/09	Entrega do Referencial Teórico e Cronograma.
28/10	Entrega da Implementação Parcial, incluindo:  Notebook do Projeto com Análise Exploratória e Préprocessamento dos dados;  Desenvolvimento do Modelo Base;  Atualização do Cronograma e do Pipeline da Solução.
18/11	Entrega Final do Projeto, abrangendo:  - Repositório no GitHub com todos os artefatos do projeto;  - Vídeos de Apresentação do Projeto e da Solução Técnica;  - Notebook do Projeto completo e executável;  - Artigo do Projeto conforme normas estabelecidas

### Referências:

- SILVA, J. L. "Impactos Ambientais do Desmatamento no Brasil: Um Estudo de Caso na Amazônia". *Revista de Ciências Ambientais*, vol. 45, 2020.
- INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE).

"Monitoramento do Desmatamento no Brasil por Satélites", 2022.

- CARVALHO, M. "Mudanças Climáticas e o Papel da Amazônia na Regulação do Clima Global". *Publicação Científica*, 2019.
- FERREIRA, A. B. "Modelagem Preditiva Aplicada ao Desmatamento na Amazônia". Tese de Doutorado, Universidade de São Paulo, 2021.



- OLIVEIRA, R. S. et al. “Aplicação de Redes Neurais na Previsão do Desmatamento na Amazônia”. *Anais da Conferência Brasileira de Inteligência Computacional*, 2019.

## **Implementação Parcial**

### **Introdução**

Nesta etapa do projeto, foi realizado a implementação parcial com foco na análise exploratória dos dados, pré-processamento e desenvolvimento de um modelo base para previsão do desmatamento no Brasil. Utilizamos dados reais referentes ao desmatamento nos estados da Amazônia Legal entre 1988 e 2022, conforme fornecido pelo INPE. Os resultados obtidos nos permitem compreender melhor os padrões históricos de desmatamento e estabelecer uma base sólida para o desenvolvimento de modelos preditivos mais robustos nas próximas etapas.

### **1. Importação das Bibliotecas Necessárias**

```
# Importando as bibliotecas necessárias
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
```

```
# Configurando a exibição de gráficos no notebook
%matplotlib inline
```

### **2. Carregamento e Visualização dos Dados**

```
# Definindo o caminho do arquivo CSV caminho_csv =  
"C:\Users\mates\Downloads\desmatamento_prodes.csv"
```

```
# Carregando o DataFrame df =  
pd.read_csv(caminho_csv)
```

```
# Visualizando as primeiras linhas do DataFrame print(df.head()) Saída:  
referencia acre amazonas amapa maranhao mato_grosso para rondonia \
```

0	1988	620	1510	60	2450	5140	6990	2340
1	1989	540	1180	130	1420	5960	5750	1430
2	1990	550	520	250	1100	4020	4890	1670
3	1991	380	980	410	670	2840	3780	1110
4	1992	400	799	36	1135	4674	3787	2265

```
roraima tocantins area_total_desmatamento
```

0	290	1650		21050
1	630	730		17770
2	150	580	13730	3 420 440 11030
4	281	409	13786	

### 3. Análise Exploratória de Dados (EDA)

#### 3.1 Informações Gerais do DataFrame

```
# Informações gerais sobre o DataFrame
```

RangeIndex: 35 entries, 0 to 34

Data columns (total 11 columns):

#	Column	Non-Null Count	Dtype
0	referencia	35 non-null	int64
1	acre	35 non-null	int64
2	amazonas	35 non-null	int64
3	amapa	35 non-null	int64
4	maranhao	35 non-null	int64
5	mato_grosso	35 non-null	int64
6	para	35 non-null	int64
7	rondonia	35 non-null	int64
8	roraima	35 non-null	int64
9	tocantins	35 non-null	int64
10	area_total_desmatamento	35 non-null	int64

dtypes: int64(11) memory usage: 3.1 KB

None

### 3.2 Verificação de Dados Faltantes

# Verificando a presença de dados faltantes

print(df.isnull().sum()) referencia

0 acre 0 amazonas

0

amapa 0

maranhao 0

mato\_grosso 0

para 0

rondonia 0

```
roraima          0
tocantins        0
area_total_desmatamento    0 dtype:
int64
```

**Análise:** Não há dados faltantes no DataFrame, o que é positivo para a qualidade da análise e modelagem.

### 3.3 Estatísticas Descritivas

```
# Estatísticas descritivas do DataFrame print(df.describe())
```

**Análise:** As estatísticas descritivas nos fornecem insights sobre a distribuição dos dados em cada estado, permitindo identificar possíveis outliers ou variações significativas entre os anos.

### 3.4 Evolução da Área Total Desmatada

```
# Gráfico da evolução da área total desmatada ao longo dos anos
plt.figure(figsize=(12,6)) sns.lineplot(data=df, x='referencia',
y='area_total_desmatamento', marker='o') plt.title('Evolução da Área Total
Desmatada (1988-2022)') plt.xlabel('Ano') plt.ylabel('Área Total Desmatada
(km²)') plt.show()
```

**Análise:** O gráfico mostra a tendência do desmatamento total ao longo dos anos. Podemos identificar períodos de aumento e diminuição, possivelmente relacionados a políticas públicas ou fatores econômicos.

### 3.5 Desmatamento por Estado

```
# Lista de estados
estados = ['acre', 'amazonas', 'amapa', 'maranhao', 'mato_grosso', 'para',
'rondonia', 'roraima', 'tocantins']
```

```
# Gráfico da evolução do desmatamento por estado

plt.figure(figsize=(14,8)) for estado in estados:

    sns.lineplot(data=df, x='referencia', y=estado, label=estado.capitalize())

plt.title('Evolução do Desmatamento por Estado (1988-2022)')

plt.xlabel('Ano') plt.ylabel('Área Desmatada (km²)')

plt.legend() plt.show()
```

**Análise:** Este gráfico nos permite comparar o desmatamento entre os estados ao longo do tempo. Observa-se que estados como Pará e Mato Grosso apresentam os maiores índices de desmatamento.

### 3.6 Matriz de Correlação

```
# Selecionando apenas as colunas dos estados df_estados

= df[estados]

# Calculando a matriz de correlação corr_matrix

= df_estados.corr()

# Exibindo o heatmap da matriz de correlação

plt.figure(figsize=(10,8)) sns.heatmap(corr_matrix,

annot=True, cmap='coolwarm') plt.title('Matriz de

Correlação entre Estados') plt.show()
```

**Análise:** A matriz de correlação nos mostra a relação entre os desmatamentos nos diferentes estados. Correlações altas podem indicar padrões similares ou fatores comuns que afetam múltiplos estados.

## 4. Pré-processamento dos Dados

## 4.1 Seleção de Variáveis

# Variáveis independentes (X) e dependente (y)

X = df[estados] y =

df['area\_total\_desmatamento']

## 4.2 Divisão em Conjuntos de Treino e Teste

# Dividindo os dados (70% treino, 30% teste)

```
X_train, X_test, y_train, y_test = train_test_split(X, y,  
                                                    test_size=0.3, random_state=42)
```

**Análise:** Dividimos os dados para avaliar a capacidade de generalização do modelo. Utilizamos 70% dos dados para treino e 30% para teste.

## 5. Desenvolvimento do Modelo Base

Aplicamos três modelos de regressão para prever a área total desmatada:

- **Regressão Linear**
- **Árvore de Decisão**
- **Random Forest**

### 5.1 Regressão Linear

# Criando o modelo de Regressão Linear

```
modelo_lr = LinearRegression()
```

```
# Treinando o modelo modelo_lr.fit(X_train,  
y_train)
```

```
# Fazendo previsões no conjunto de teste y_pred_lr  
= modelo_lr.predict(X_test)
```

```
# Avaliando o modelo
```

```
mae_lr = mean_absolute_error(y_test, y_pred_lr)
```

```
mse_lr = mean_squared_error(y_test, y_pred_lr)
```

```
rmse_lr = np.sqrt(mse_lr) r2_lr =
```

```
r2_score(y_test, y_pred_lr)
```

```
print('Desempenho do Modelo de Regressão Linear:')
```

```
print(f'MAE: {mae_lr:.2f}') print(f'RMSE: {rmse_lr:.2f}')
```

```
print(f'R²: {r2_lr:.2f}')
```

Desempenho do Modelo de Regressão Linear:

MAE: 0.00

RMSE: 0.00

R²: 1.00

**Análise:** O modelo de Regressão Linear apresentou um desempenho aparentemente perfeito, com erro zero e  $R^2$  igual a 1. Isso indica que o modelo está superajustado (overfitting) aos dados de treino, possivelmente devido a alguma inconsistência nos dados ou no processo de modelagem.

## 5.2 Árvore de Decisão

```
# Criando o modelo de Árvore de Decisão modelo_dt =
```

```
DecisionTreeRegressor(random_state=42)
```

```
# Treinando o modelo modelo_dt.fit(X_train,
```

```
y_train)
```

```
# Fazendo previsões no conjunto de teste y_pred_dt
= modelo_dt.predict(X_test)

# Avaliando o modelo
mae_dt = mean_absolute_error(y_test, y_pred_dt)
mse_dt = mean_squared_error(y_test, y_pred_dt)
rmse_dt = np.sqrt(mse_dt) r2_dt =
r2_score(y_test, y_pred_dt)

print('\nDesempenho do Modelo de Árvore de Decisão:')
print(f'MAE: {mae_dt:.2f}') print(f'RMSE: {rmse_dt:.2f}')
print(f'R²: {r2_dt:.2f}')
```

Desempenho do Modelo de Árvore de Decisão:

MAE: 2083.55

RMSE: 2625.86

R²: 0.88

**Análise:** A Árvore de Decisão apresentou um desempenho razoável, com R² de 0.88. No entanto, o erro médio absoluto (MAE) e o RMSE são relativamente altos, indicando que o modelo não está prevendo com alta precisão.

### 5.3 Random Forest

```
# Criando o modelo de Random Forest modelo_rf =
RandomForestRegressor(n_estimators=100, random_state=42)
```



```

# Treinando o modelo modelo_rf.fit(X_train,
y_train)

# Fazendo previsões no conjunto de teste y_pred_rf
= modelo_rf.predict(X_test)

# Avaliando o modelo

mae_rf = mean_absolute_error(y_test, y_pred_rf)
mse_rf = mean_squared_error(y_test, y_pred_rf)
rmse_rf = np.sqrt(mse_rf) r2_rf =
r2_score(y_test, y_pred_rf)

print('\nDesempenho do Modelo de Random Forest:')
print(f'MAE: {mae_rf:.2f}') print(f'RMSE:
{rmse_rf:.2f}') print(f'R²: {r2_rf:.2f}')

```

Desempenho do Modelo de Random Forest:

MAE: 1640.12

RMSE: 2294.24

R²: 0.91

**Análise:** O modelo de Random Forest apresentou o melhor desempenho entre os modelos testados, com R² de 0.91 e menores valores de MAE e RMSE em comparação com a Árvore de Decisão.

## 5.4 Comparação dos Modelos

```

# Criando um DataFrame para comparar as métricas resultados
= pd.DataFrame({

```

```

'Modelo': ['Regressão Linear', 'Árvore de Decisão', 'Random Forest'],
'MAE': [mae_lr, mae_dt, mae_rf],
'RMSE': [rmse_lr, rmse_dt, rmse_rf],
'R²': [r2_lr, r2_dt, r2_rf]
})

```

```
print("\nComparação dos Modelos:") print(resultados)
```

Comparação dos Modelos:

	Modelo	MAE	RMSE	R²
0	Regressão Linear	2.397759e-12	3.627629e-12	1.000000
1	Árvore de Decisão	2.083545e+03	2.625860e+03	0.881437
2	Random Forest	1.640119e+03	2.294237e+03	0.909493

**Análise:** A tabela resume o desempenho dos modelos. Apesar da Regressão Linear apresentar métricas perfeitas, isso é suspeito e indica overfitting. O Random Forest mostrou-se superior à Árvore de Decisão, com melhores métricas de erro e R² mais alto.

## 5.5 Visualização das Previsões

```

# Criando um DataFrame com as previsões e valores reais df_resultados
= pd.DataFrame({
    'Valor Real': y_test.reset_index(drop=True),
    'Regressão Linear': y_pred_lr,
    'Árvore de Decisão': y_pred_dt,
    'Random Forest': y_pred_rf
})

```

```
# Plotando os resultados plt.figure(figsize=(12,6)) plt.plot(df_resultados['Valor Real'], label='Valor Real', marker='o') plt.plot(df_resultados['Regressão Linear'], label='Regressão Linear', marker='x') plt.plot(df_resultados['Árvore de Decisão'], label='Árvore de Decisão', marker='^') plt.plot(df_resultados['Random Forest'], label='Random Forest', marker='s') plt.title('Comparação das Previsões dos Modelos') plt.xlabel('Amostras') plt.ylabel('Área Total Desmatada (km²)') plt.legend() plt.show()
```

**Análise:** O gráfico permite visualizar a proximidade das previsões de cada modelo em relação aos valores reais. Observa-se que o Random Forest acompanha melhor a tendência dos dados reais.

## 6. Discussão dos Resultados A análise inicial sugere

que:

- **Regressão Linear:** O desempenho perfeito indica um possível overfitting ou erro nos dados. É improvável que um modelo linear simples consiga capturar perfeitamente a complexidade do desmatamento.
- **Árvore de Decisão:** Embora capture relações não lineares, pode sofrer de overfitting e não generalizar bem para novos dados.
- **Random Forest:** Apresentou o melhor desempenho, reduzindo o overfitting e melhorando a capacidade de generalização.

### Possíveis Motivos para o Overfitting na Regressão Linear:

- **Colinearidade Perfeita:** Se a soma das áreas desmatadas nos estados for exatamente igual à área total desmatada, o modelo pode ajustar perfeitamente os dados.
- **Dados Limitados:** Com apenas 35 amostras, é difícil para o modelo generalizar, especialmente se a divisão treino-teste não for representativa.

### Soluções:

- **Adicionar Variáveis Externas:** Incluir dados socioeconômicos e ambientais para enriquecer o modelo.
- **Aumentar o Conjunto de Dados:** Se possível, obter dados de mais anos ou de outras regiões.

## 7. Próximos Passos

- **Feature Engineering:** Criar novas variáveis que possam melhorar o modelo.
- **Otimização de Hiperparâmetros:** Utilizar Grid Search ou Random Search para encontrar os melhores parâmetros dos modelos.
- **Validação Cruzada:** Implementar K-Fold Cross Validation para avaliar a capacidade de generalização dos modelos.
- **Incluir Novos Modelos:** Testar modelos avançados como XGBoost e LightGBM.

## 9. Atualização do Pipeline Proposto

- **Coleta de Novos Dados:** Buscar dados adicionais que possam enriquecer o modelo.
- **Análise Exploratória Aprofundada:** Realizar análises mais detalhadas para identificar padrões ocultos.
- **Modelagem Avançada:** Implementar modelos mais sofisticados e técnicas de ensemble.
- **Avaliação e Interpretação:** Focar na interpretabilidade dos modelos para apoiar a tomada de decisão.

## 10. Conclusão

A implementação parcial deste projeto proporcionou insights significativos sobre a dinâmica do desmatamento na Amazônia Legal brasileira ao longo dos últimos 35 anos. Através da análise exploratória dos dados reais fornecidos pelo INPE, identificamos padrões temporais e espaciais que refletem as complexas interações entre fatores socioeconômicos, ambientais e políticos que influenciam o desmatamento.

A aplicação de modelos de regressão linear, árvore de decisão e random forest permitiu avaliar diferentes abordagens para a previsão da área total desmatada. Os resultados indicaram que:

- **Regressão Linear:** Embora tenha apresentado métricas perfeitas (MAE e RMSE próximos de zero e  $R^2$  igual a 1), esse desempenho é irrealista e sugere um overfitting extremo ou possível multicolinearidade perfeita nos dados. Isso ocorre possivelmente porque a variável dependente (área total desmatada) é a soma direta das variáveis independentes (desmatamento por estado), resultando em uma relação linear perfeita. Portanto, este modelo não é adequado para previsões generalizáveis neste contexto.
- **Árvore de Decisão:** Apresentou um desempenho razoável, com  $R^2$  de 0,88, mas com erros médios elevados. Isso indica que o modelo captura algumas das relações não lineares entre as variáveis, mas ainda carece de capacidade preditiva robusta, possivelmente devido à alta variabilidade dos dados e à sensibilidade do modelo a overfitting.
- **Random Forest:** Obteve o melhor desempenho entre os modelos testados, com  $R^2$  de 0,91 e menores valores de MAE e RMSE em comparação com a árvore de decisão. Isso sugere que a random forest, ao combinar múltiplas árvores de decisão e reduzir a variância, é mais eficaz em capturar padrões complexos nos dados e oferecer previsões mais precisas.

### Análise Crítica dos Resultados:

A análise dos resultados indica que, embora o modelo de random forest seja promissor, ainda existem desafios a serem superados para melhorar a precisão das previsões. Alguns pontos críticos incluem:

- **Limitação dos Dados:** O conjunto de dados possui apenas 35 observações, o que é relativamente pequeno para treinamento de modelos de aprendizado de máquina, especialmente aqueles que exigem grande quantidade de dados para generalizar efetivamente.
- **Colinearidade das Variáveis:** A soma das áreas desmatadas por estado resulta na área total desmatada, o que cria uma relação linear perfeita entre as variáveis independentes e a dependente. Isso limita a capacidade dos modelos de identificar padrões mais sutis e dependências não lineares.
- **Ausência de Variáveis Externas:** O modelo atual considera apenas o desmatamento por estado como variável preditora. Não são consideradas outras variáveis que influenciam o desmatamento, como fatores econômicos (preço de commodities, PIB regional), sociais (Índices de

pobreza, migração), ambientais (índices de seca, precipitação) e políticos (mudanças nas políticas ambientais, fiscalização).

### **Implicações e Recomendações:**

Para aprimorar o modelo preditivo e tornar as previsões mais confiáveis e úteis para a formulação de políticas públicas, recomendamos:

1. **Incorporação de Variáveis Externas:** A inclusão de dados socioeconômicos, ambientais e políticos pode enriquecer o modelo, permitindo capturar os fatores subjacentes que impulsionam o desmatamento. Isso pode incluir variáveis como:
  - Taxa de crescimento populacional.
  - Índices de desenvolvimento humano (IDH).
  - Preços internacionais de soja e carne bovina.
  - Dados de precipitação e temperatura.
  - Mudanças na legislação ambiental.
2. **Aumento do Conjunto de Dados:** Expandir o período de análise ou incluir dados de outras regiões pode aumentar o número de observações e melhorar a capacidade dos modelos de generalizar.
3. **Feature Engineering Avançado:** Criar novas features que possam capturar tendências temporais (como médias móveis), variações sazonais ou interações entre estados.
4. **Otimização e Validação do Modelo:**
  - **Validação Cruzada:** Implementar K-Fold Cross Validation para avaliar a robustez do modelo e evitar overfitting.
  - **Otimização de Hiperparâmetros:** Utilizar técnicas como Grid Search ou Random Search para encontrar os melhores parâmetros dos modelos.
5. **Interpretação e Transparência:** Utilizar métodos de interpretabilidade de modelos (como SHAP values) para entender o impacto de cada variável nas previsões e fornecer insights acionáveis para os tomadores de decisão.

### **Contribuições do Estudo:**

Este projeto contribui para a compreensão do desmatamento no Brasil ao:

- Fornecer uma análise histórica detalhada do desmatamento por estado na Amazônia Legal.
- Demonstrar a aplicação de diferentes técnicas de aprendizado de máquina na modelagem de fenômenos ambientais complexos.
- Identificar as limitações dos modelos quando aplicados a dados ambientais e socioeconômicos, destacando a importância de uma abordagem multidisciplinar.

### **Reflexão Final:**

O desmatamento é um problema multifacetado que exige soluções integradas e baseadas em evidências. Embora os modelos preditivos sejam ferramentas valiosas, eles devem ser complementados com conhecimento local, participação das comunidades afetadas e comprometimento político. Este projeto busca não apenas desenvolver um modelo matemático, mas também contribuir para a conscientização e a elaboração de estratégias eficazes de conservação e desenvolvimento sustentável.

Ao avançarmos com este estudo, esperamos fornecer insights que possam apoiar a mitigação dos impactos do desmatamento, preservar a biodiversidade e promover o equilíbrio climático global, alinhando-se aos objetivos de desenvolvimento sustentável e aos acordos internacionais de proteção ambiental.