

# Relatório Final: Modelagem Preditiva e Recomendações Estratégicas para os Impactos do Desmatamento no Pará

## Zetta Lab 2025 - Desafio II: Ciência e Governança de Dados

**Autor(a):** Breno Carvalho Pedroso



Repositório do Projeto (contém Notebooks, Dados, Modelos e Dashboard):

[https://github.com/Boakpe/ZettaLab2025\\_Dasafio1\\_Dados](https://github.com/Boakpe/ZettaLab2025_Dasafio1_Dados)

## 1. Introdução

Este relatório documenta a segunda e conclusiva fase do projeto, que avança da exploração para a predição dos impactos do desmatamento no estado do Pará. O objetivo central é responder à pergunta: "Como poderíamos avaliar e prever os impactos socioeconômicos do desmatamento no estado do Pará?". Para isso, foram desenvolvidos modelos de Machine Learning para antecipar tendências em indicadores-chave.

O resultado mais significativo e contraintuitivo desta fase foi a revelação de uma desconexão entre o ritmo do desmatamento e a evolução dos principais indicadores socioeconômicos municipais. Ao contrário da narrativa comum, os dados e os modelos não suportam uma correlação direta e imediata entre o aumento da derrubada da floresta e uma melhoria tangível no PIB per capita ou uma diminuição na dependência de benefícios sociais. Esta aparente contradição tornou-se o eixo central da análise, direcionando o trabalho para entender as complexas razões por trás dessa dinâmica e formulando recomendações baseadas nesta realidade mais nuançada.

## 2. Metodologia de Modelagem Preditiva

Para construir um sistema de previsão robusto e investigar essas relações complexas, foi adotada uma metodologia estruturada, desde a seleção de dados e engenharia de features até a validação e otimização dos modelos.

### 2.1. Escolha da Abordagem e Inclusão de Novos Dados

Nesta segunda fase, uma nova e crucial variável foi incorporada ao conjunto de dados: o número de **Internações por Doenças Respiratórias**, obtido do DATASUS. A análise exploratória da primeira fase revelou a ausência de um indicador direto do impacto na saúde pública. A inclusão desta variável foi fundamental para quantificar a relação entre as queimadas (frequentemente associadas ao desmatamento) e a saúde da população.

A abordagem final consistiu em treinar quatro modelos distintos para prever a evolução anual das seguintes variáveis-alvo:

1. **PIB per capita (R\$):** Indicador de bem-estar econômico.
2. **VAB Agropecuária (R\$ 1.000):** Medida da atividade do setor primário.
3. **Total de Benefícios Básicos (Bolsa Família):** Proxy para vulnerabilidade social.
4. **Internações por Doenças Respiratórias:** Indicador do impacto na saúde.

### 2.2. Engenharia de Features

Para que os modelos pudessem aprender com a dinâmica temporal dos dados, foi implementada uma extensa engenharia de features. A função `criar_features` gerou variáveis cruciais a partir dos dados brutos:

- **Lags Temporais (Shifts):** Valores de features de anos anteriores (e.g., `Desmatamento_lag1`, `PIB_lag1`).
- **Taxas de Crescimento (Growth):** Variação percentual em relação ao ano anterior.
- **Médias Móveis (Rolling Means):** Médias de janelas de 2 e 3 anos para suavizar tendências.
- **Tendência Linear (Trend):** Inclinação da reta de regressão dos últimos 3 anos.

Essa engenharia transforma um conjunto de séries temporais (dados em painel) em um formato de dados tabulares, permitindo que algoritmos de aprendizado de máquina tradicionais possam ser aplicados de forma eficiente, aprendendo tanto com os valores absolutos quanto com as dinâmicas e tendências dos dados.

### 2.3. Escolha do Modelo: LightGBM

O algoritmo **LightGBM (Light Gradient Boosting Machine)** foi selecionado para todos os modelos, justificado por suas vantagens:

- **Alto Desempenho:** Reconhecido por sua precisão em dados tabulares.
- **Eficiência Computacional:** Crucial para a otimização de hiperparâmetros.
- **Flexibilidade:** Lida nativamente com valores ausentes e variáveis categóricas (como `Município`).
- **Estado da Arte em Séries Temporais:** LightGBM tem se destacado como uma das melhores opções para previsão em séries temporais quando combinado com engenharia de features adequada, devido à sua capacidade de capturar padrões complexos e não lineares. [Artigo de referência](#).

### 2.4. Validação, Otimização e Desempenho

Foi implementado um rigoroso processo de validação temporal para evitar vazamento de dados (*data leakage*):

- **Treino:** Dados até 2020.
- **Validação:** Dados de 2021 (para otimização de hiperparâmetros com `Optuna`).
- **Teste:** Dados de 2022 (para avaliação final do modelo).

O desempenho final dos modelos no conjunto de teste (ano de 2022) demonstrou excelente poder preditivo:

Modelo	Métrica	Valor	Descrição
PIB per Capita	R <sup>2</sup>	0.941	Coeficiente de Determinação
VAB Agropecuária	R <sup>2</sup>	0.985	Coeficiente de Determinação
Benefícios Sociais	R <sup>2</sup>	0.989	Coeficiente de Determinação
Saúde Respiratória	R <sup>2</sup>	0.887	Coeficiente de Determinação

Os altos valores de R<sup>2</sup> indicam que os modelos foram tecnicamente bem-sucedidos em explicar a variância dos dados. A questão fundamental, portanto, não é a capacidade preditiva do modelo, mas *o que ele nos revela sobre a importância relativa de cada fator*.

## 3. Análise de Resultados: A Desconexão Revelada

A descoberta mais importante do estudo veio da análise de importância das features (SHAP e LightGBM Gain). Para os modelos de **PIB per capita** e **Benefícios Sociais**, o **Desmatamento (km²)\_lag1** não figurou entre os principais preditores. O fator consistentemente mais importante foi:

- **A própria inércia histórica da variável** (ex: **PIB per capita\_lag1**).

A análise SHAP, embora detecte um *signal estatístico* fraco e por vezes positivo entre **VAB Agropecuária\_lag1/Desmatamento\_lag1** e o PIB, confirma que este impacto é **marginal** quando comparado aos outros fatores. Isso explica por que, ao visualizar os dados no dashboard, picos de desmatamento (como os observados entre 2019-2022) não se traduzem em saltos proporcionais nos indicadores socioeconômicos.

#### Hipóteses para esta Desconexão:

- **Economia Ilegal e Especulativa:** Grande parte do desmatamento pode ser impulsionada pela grilagem e extração ilegal de recursos, cujos lucros não circulam na economia formal do município e, portanto, não impactam o PIB per capita ou reduzem a pobreza.
- **Concentração de Renda:** Os benefícios da agropecuária expansionista podem estar concentrados em poucos grandes proprietários ou corporações, sem gerar uma distribuição de renda que altere significativamente a média municipal.
- **Defasagem Temporal:** Os impactos econômicos podem ocorrer em escalas de tempo mais longas do que as capturadas na análise anual.

**A Exceção que Confirma a Regra: Saúde Pública** Em nítido contraste, a análise do modelo de saúde revelou uma **relação direta, forte e inequívoca** entre **Focos de Queimada\_lag1** e o **aumento de Internações por Doenças Respiratórias**. Este é um impacto físico, imediato e mensurável, provando que os modelos são capazes de detectar correlações fortes quando elas de fato existem nos dados.

## 4. Dashboard Interativo como Ferramenta de Investigação

Com base nessas descobertas, o dashboard interativo foi concebido não apenas como uma ferramenta de previsão, mas como uma **plataforma de análise investigativa**. Ele permite que gestores e pesquisadores explorem visualmente essa desconexão: um usuário pode selecionar um município, observar um pico de desmatamento e verificar, nas outras abas, a ausência de uma resposta socioeconômica correspondente. A ferramenta, assim, capacita a desconstrução de narrativas simplistas que ligam o desmatamento diretamente ao desenvolvimento.

## 5. Recomendações Estratégicas

As recomendações foram formuladas para enfrentar o problema central identificado: um modelo de exploração ambiental que não gera prosperidade local distribuída.

### 1. Foco em Investigar e Combater a Economia Ilegal Associada ao Desmatamento:

- **Insight:** O aumento do desmatamento não se correlaciona com a melhoria do PIB per capita local, sugerindo atividades ilícitas ou especulativas.
- **Recomendação:** Utilizar o cruzamento de dados para identificar "zonas de alerta": municípios com **alto desmatamento e baixo crescimento econômico**. Essas áreas devem ser priorizadas para ações de inteligência e fiscalização focadas em coibir a grilagem e a extração ilegal de recursos.

## 2. Implementação de um Sistema de Alerta e Prevenção de Saúde Pública:

- **Insight:** A ligação entre queimadas e doenças respiratórias é clara, direta e preditível.
- **Recomendação:** Integrar os dados do INPE e as previsões do modelo com a Secretaria de Saúde do Estado para alocar proativamente recursos (leitos, medicamentos) e emitir alertas para a população em municípios sob alto risco.

## 3. Desenvolvimento de Políticas de Bioeconomia como Alternativa Real:

- **Insight:** O modelo extrativista atual, impulsionado pelo desmatamento, falha em gerar desenvolvimento socioeconômico inclusivo.
- **Recomendação:** Intensificar os investimentos em cadeias produtivas da sociobiodiversidade (castanha, açaí, etc.), ecoturismo e sistemas agroflorestais, priorizando os municípios mais dependentes do VAB Agropecuário e com alta pressão de desmatamento.

## 4. Fortalecimento da Rede de Proteção Social como Política Estrutural:

- **Insight:** A vulnerabilidade social persiste independentemente dos ciclos de desmatamento, pois os benefícios econômicos da exploração não chegam à base da população.
- **Recomendação:** Tratar os programas de transferência de renda como uma política estrutural indispensável para garantir a dignidade em uma região onde o modelo de "desenvolvimento" predominante não é inclusivo.

## 6. Conclusão

Este projeto partiu da tarefa de prever os impactos do desmatamento e chegou a uma conclusão mais profunda: a **evidência quantitativa da desconexão entre a degradação ambiental e a prosperidade local no Pará**. Os modelos de Machine Learning, ao mesmo tempo que demonstraram alta precisão técnica, revelaram que a trajetória socioeconômica de um município é muito mais ditada por sua própria inércia histórica e características estruturais do que pelo ritmo anual da destruição florestal.

A principal contribuição deste trabalho é o fornecimento de uma base de dados e ferramentas que desafiam a narrativa simplista do "desmatar para desenvolver". As descobertas clamam por uma mudança de paradigma nas políticas públicas, direcionando o foco para o combate às economias ilícitas e para a promoção de um modelo de bioeconomia que seja verdadeiramente capaz de gerar desenvolvimento sustentável e inclusivo para as populações da Amazônia.