

Ano V, v.2 2025 | submissão: 05/12/2025 | aceito: 08/12/2025 | publicação: 10/12/2025

Análise de Dados Socioeconômicos do Vale do Sepotuba Para Desenvolvimento de um Modelo Preditivo de Crescimento Econômico

Analysis of Socioeconomic Data From the Sepotuba Valley for the Development of a Predictive Model of Economic Growth

Luiz Henrique Camargo Moura¹ – Universidade do Estado de Mato Grosso Prof. Me. Leandro Mazurek² – Universidade do Estado de Mato Grosso

Prof.^a Ma. Raquel da Silva Vieira Coelho³ – Universidade do Estado de Mato Grosso

Resumo

Este artigo propõe a análise de dados socioeconômicos do Vale do Sepotuba com o objetivo de desenvolver um modelo preditivo de crescimento econômico baseado em técnicas de aprendizado de máquina. A pesquisa adota uma abordagem quantitativa e descritiva, integrando fundamentos de ciência de dados, economia regional e desenvolvimento sustentável. O estudo busca compreender os fatores que influenciam o crescimento regional, estruturando dados históricos e atuais de indicadores como PIB, escolaridade, desemprego e infraestrutura. A partir da modelagem preditiva, pretende-se gerar insights que possam subsidiar políticas públicas e decisões estratégicas para a região.

Palavras-chave: Aprendizado de Máquina; Crescimento Econômico; Análise Socioeconômica; Vale do Sepotuba.

Abstract

This article proposes the analysis of socioeconomic data from the Vale do Sepotuba in order to develop a predictive economic growth model based on machine learning techniques. The research adopts a quantitative and descriptive approach, integrating data science, regional economics, and sustainable development. The study seeks to understand the factors that influence regional growth, structuring historical and current data of indicators such as GDP, education, unemployment, and infrastructure. Through predictive modeling, the aim is to generate insights that can support public policies and strategic decisions for the region.

Keywords: Machine Learning; Economic Growth; Socioeconomic Analysis; Vale do Sepotuba.

1 . INTRODUÇÃO

O crescimento econômico regional é um fenômeno complexo, influenciado por múltiplos fatores interdependentes, como capital humano, infraestrutura, inovação tecnológica, educação e políticas públicas. No contexto do Vale do Sepotuba, no sudoeste de Mato Grosso, compreender essas dinâmicas torna-se ainda mais desafiador devido às particularidades históricas, territoriais e socioeconômicas que moldaram o desenvolvimento local. A região reúne municípios de destaque na economia mato-grossense, como Tangará da Serra, Campo Novo do Parecis e Sapezal, ao mesmo tempo em que apresenta localidades com limitações estruturais, como Lambari D'Oeste, Salto do Céu e Porto Estrela. Essa diversidade.

interna reflete uma realidade marcada por contrastes, com polos agroindustriais integrados a cadeias globais de commodities coexistindo com regiões de baixa diversificação produtiva e menor acesso a serviços públicos essenciais.

Apesar do potencial produtivo, o desenvolvimento econômico da região enfrenta barreiras significativas, incluindo desigualdade territorial, infraestrutura insuficiente em alguns municípios e

Ano V, v.2 2025 | submissão: 05/12/2025 | aceito: 08/12/2025 | publicação: 10/12/2025

fragmentação de dados socioeconômicos, o que dificulta diagnósticos precisos e a formulação de políticas públicas eficazes. De acordo com dados recentes do IBGE (2023) e do IPEA (2023), o PIB per capita no Vale do Sepotuba varia entre R\$ 28.000 em municípios menores e R\$ 55.000 nos polos agroindustriais; o Índice de Desenvolvimento Humano Municipal apresenta variações de 0,65 a 0,78; e a taxa de desemprego regional oscila entre 5% e 10%, dependendo do município. Esses indicadores evidenciam desigualdades internas e reforçam a necessidade de análises detalhadas que considerem o contexto específico de cada localidade.

Nesse cenário, o uso de técnicas de aprendizado de máquina surge como uma alternativa promissora, uma vez que, diferentemente de métodos estatísticos tradicionais, esses algoritmos são capazes de identificar padrões ocultos em grandes volumes de dados, explorar relações não lineares entre variáveis e gerar previsões robustas em cenários incertos. Aplicados ao contexto regional, podem revelar determinantes do crescimento econômico que não seriam facilmente perceptíveis por meio de análises convencionais. Estudos recentes destacam que a integração entre análise de dados, Big Data e aprendizado de máquina pode auxiliar gestores públicos e investidores privados a compreender melhor as dinâmicas de crescimento, direcionando recursos para áreas estratégicas e reduzindo riscos de investimento (Mayer-Schönberger; Cukier, 2013; Provost; Fawcett, 2013). No Vale do Sepotuba, essa aplicação não apenas fortalece a competitividade agroindustrial, mas também fomenta um desenvolvimento equilibrado, contemplando educação, infraestrutura, sustentabilidade ambiental e inovação tecnológica.

Durante a revisão bibliográfica, não foram encontrados estudos que aplicassem modelos de aprendizado de máquina especificamente aos municípios do Vale do Sepotuba. Essa lacuna evidencia a originalidade da pesquisa e reforça sua contribuição científica, ao explorar um contexto regional ainda pouco estudado com técnicas de análise preditivas avançadas.

Diante desse contexto, esta pesquisa busca responder à seguinte questão: como aplicar técnicas de aprendizado de máquina para prever o crescimento econômico dos municípios do Vale do Sepotuba a partir de dados socioeconômicos disponíveis, de forma a subsidiar políticas públicas e decisões estratégicas do setor privado? Para atingir esse objetivo, o estudo propõe coletar e organizar dados socioeconômicos históricos e atuais, aplicar modelos de previsão, como Regressão Linear e Random Forest, avaliar o desempenho desses modelos e interpretar os resultados obtidos, gerando recomendações aplicáveis tanto ao setor público quanto ao privado. Assim, o objetivo geral consiste em propor e analisar um modelo preditivo de crescimento econômico para os municípios da região, enquanto os objetivos específicos incluem reunir dados confiáveis, aplicar técnicas de aprendizado de máquina para análise preditiva, avaliar a precisão dos modelos e elaborar recomendações fundamentadas para políticas públicas e estratégias de investimento.

A relevância do estudo está em fornecer subsídios para um planejamento municipal

Ano V, v.2 2025 | submissão: 05/12/2025 | aceito: 08/12/2025 | publicação: 10/12/2025

mais eficiente, apoiar políticas públicas e reduzir riscos para o setor privado. A inovação reside na aplicação de aprendizado de máquina para análise preditiva em contexto municipal permitindo identificar determinantes do crescimento econômico que métodos tradicionais poderiam não revelar, contribuindo para um desenvolvimento regional mais justo, sustentável e inovador.

O artigo está estruturado da seguinte forma: o Capítulo 2 apresenta a revisão de literatura sobre crescimento econômico regional, análise de dados e aprendizado de máquina o Capítulo 3 detalha a metodologia adotada, incluindo a coleta de dados e a aplicação dos modelos preditivos; o Capítulo 4 discute os resultados obtidos e suas implicações; e o Capítulo 5 apresenta as conclusões e recomendações para políticas públicas e futuras pesquisas.

2 . FUNDAMENTAÇÃO TEÓRICA

Ao tratar do crescimento econômico, especialmente em regiões que enfrentam desafios significativos, como o Vale do Sepotuba, é insuficiente limitar a análise a dados quantitativos ou à simples aplicação de teorias convencionais. Torna-se essencial considerar o contexto social e regional, interpretando os dados de forma crítica e compreendendo as dinâmicas que ocorrem no território. O Vale do Sepotuba compreende diversos municípios do sudoeste mato-grossense, incluindo Tangará da Serra, Campo Novo do Parecis, Sapezal Barra do Bugres, Denise, Nova Olímpia, Lambari D'Oeste, Rio Branco, Salto do Céu e Porto Estrela (SEMA-MT, 2021). Diante disso, esta etapa do estudo integra a análise de dados, técnicas de aprendizado de máquina e os fundamentos da economia regional, com o objetivo de construir uma base analítica robusta que contribua para a projeção do crescimento econômico da região.

2.1 Análise de dados, aprendizado de máquina e Big Data no apoio ao desenvolvimento regional

Analisar dados socioeconômicos vai além de identificar padrões econômicos, é também compreender as relações sociais mais complexas. Indicadores como o Produto Interno Bruto (PIB), a taxa de desemprego, o nível médio de escolaridade, os investimentos e a infraestrutura mostram realidades que impactam diretamente o desenvolvimento da região do Vale do Sepotuba. Esses dados não servem apenas para preencher planilhas, mas são ferramentas poderosas para transformar realidades. Com base neles, é possível pensar políticas públicas mais justas e tomar decisões estratégicas que realmente melhorem a vida das pessoas na região.

Como lembra Wooldridge (2019), ao usar a modelagem estatística dentro da econometria, não

Ano V, v.2 2025 | submissão: 05/12/2025 | aceito: 08/12/2025 | publicação: 10/12/2025

estamos apenas identificando relações entre números, mas entendendo histórias contadas pelos dados, que nos ajudam a prever cenários futuros e agir com mais precisão. Ao cruzar essas informações, percebemos como os fatores socioeconômicos se entrelaçam e influenciam o crescimento da região, abrindo espaço para ações mais bem direcionadas.

O aprendizado de máquina (machine learning) vem ganhando cada vez mais espaço na previsão econômica devido à sua capacidade de lidar com grandes volumes de dados, encontrar padrões complexos e gerar previsões baseadas em evidências sólidas. Com o uso de algoritmos supervisionados, como regressão linear, redes neurais e árvores de decisão, é possível criar modelos precisos mesmo em cenários com múltiplas variáveis, o que é especialmente útil em regiões com características específicas, como o Vale do Sepotuba. Segundo Hastie, Tibshirani e Friedman (2009), o diferencial do aprendizado de máquina está na habilidade de identificar relações não lineares e sutis, que métodos tradicionais deixam escapar.

A era do Big Data ampliou a capacidade de coleta e análise de informações, mudando a forma de pensar e formular políticas públicas. Segundo Kitchin (2014), a ciência baseada em dados permite gerar insights que ultrapassam abordagens tradicionais, tornando os processos decisórios mais ágeis e responsivos. No contexto municipal e regional, a análise preditiva orientada por dados se torna uma ferramenta essencial para a otimização de políticas públicas, oferecendo previsões mais fundamentadas e eficientes.

Portanto, a integração entre análise de dados, aprendizado de máquina e Big Data representa um avanço importante no planejamento do desenvolvimento regional, permitindo decisões baseadas em evidências e fortalecendo a eficiência da gestão pública.

2.2 Crescimento econômico e tecnologia como vetores do desenvolvimento humano

O crescimento econômico de uma região vai além do aumento do PIB. É necessário observar como ele se reflete na vida das pessoas, no acesso à educação, saúde, infraestrutura e emprego. A teoria do crescimento endógeno, proposta por Romer (1990), destaca o papel do conhecimento, da inovação e do capital humano como pilares centrais para um desenvolvimento sustentável.

Investir em análise de dados e modelos preditivos é, portanto, uma estratégia essencial para o planejamento regional. Isso possibilita identificar fatores que impulsionam ou dificultam o crescimento e orientar políticas públicas e investimentos privados mais eficientes.

Contudo, é fundamental compreender que o uso da tecnologia não deve estar dissociado dos valores humanos. O uso ético da inteligência artificial, a transparência dos dados públicos e o respeito à diversidade cultural são princípios essenciais para um desenvolvimento justo e sustentável. Como destaca Diniz (2013), a tecnologia deve servir à equidade e à justiça social, e não apenas à eficiência

2.3 Aspectos históricos, modernização agrícola e desafios territoriais no Vale do Sepotuba

A análise do crescimento econômico no Vale do Sepotuba requer um olhar histórico sobre os ciclos de ocupação e o uso dos recursos naturais. A ocupação do sudoeste mato-grossense intensificou-se na década de 1970 com políticas federais de integração, como o Programa de Integração Nacional (PIN), que incentivaram a migração e a expansão da fronteira agrícola. Segundo Oliveira (2007), esse processo consolidou um modelo de desenvolvimento baseado na agroindústria exportadora, com baixa diversificação produtiva.

Durante as décadas seguintes, municípios como Sapezal, Campo Novo do Parecis e Tangará da Serra receberam grandes investimentos privados em infraestrutura rural e mecanização, enquanto outras áreas, como Lambari D'Oeste, Salto do Céu e Porto Estrela, permaneceram com economias de menor escala e dependência agropecuária (SILVA et al., 2022; LEANDRO, 2020). Essa diferenciação ainda se reflete nos indicadores socioeconômicos atuais.

A modernização agrícola aumentou a produtividade e inseriu a região na cadeia global de exportação, mas também gerou impactos sociais, como concentração fundiária e deslocamento populacional. Conforme Delgado (2012), o avanço tecnológico no campo provocou a exclusão de trabalhadores e reforçou desigualdades. Municípios com melhor infraestrutura se desenvolveram rapidamente, enquanto outros permaneceram à margem evidenciando o paradoxo da produtividade (BRANDÃO, 2007).

Os gargalos logísticos e de infraestrutura permanecem entre os principais desafios. De acordo com o IMEA (2021), a falta de rodovias asfaltadas e conectividade digital afeta diretamente a competitividade e perpetua desigualdades. Superar essas limitações requer não apenas obras físicas, mas também o uso estratégico de dados para mapear cadeias produtivas, identificar vocações locais e prever retornos de investimento social.

2.4 Educação, inovação e sustentabilidade regional

O desenvolvimento econômico sustentável está fortemente relacionado à qualidade da educação e à valorização do capital humano. No Vale do Sepotuba, onde a economia é predominantemente agropecuária, a formação técnica e científica é essencial para diversificação e inovação. Conforme Romer (1990), o conhecimento e a capacidade de inovação são motores do crescimento de longo prazo.

Municípios que investem na formação de seus jovens e estimulam parcerias entre escolas técn

Ano V, v.2 2025 | submissão: 05/12/2025 | aceito: 08/12/2025 | publicação: 10/12/2025

cas, un vers dades e setor produt vo cr am amb entes ma s favoráve s à novação. No contexto regional, o fortalecimento da UNEMAT, dos institutos federais e de centros de inovação rural pode ajudar a transição de uma economia baseada em recursos naturais para uma baseada em conhec mento.

De acordo com Sachs (2004) e Veiga (2008), o desenvolvimento sustentável precisa integrar crescimento econômico, equidade social e preservação ambiental. No Vale do Sepotuba, isso significa equilibrar o avanço agroindustrial com o apoio à agricultura familiar à educação técnica e à conservação ambiental, construindo um modelo regiona verdade ramente sustentável.

2.5 Tecnologia, governo digital e análise espacial para o planejamento regional

Com o avanço da transformação digital, o uso de tecnologias no setor público tornou-se essencial para uma gestão ma s ef c ente e transparente. Ferramentas como dashboards sistemas de informação geográfica (SIG), inteligência artificial e plataformas colaborativas ampliam a capacidade de planejamento e acompanhamento de indicadores socioeconômicos.

No Vale do Sepotuba, o governo digital pode melhorar a alocação de recursos, o monitoramento de políticas públicas e a prestação de serviços essenciais, além de fortalecer a participação cidadã e a governança democrática. Castells (2000) destaca que o uso inteligente das tecnologias de informação é crucial para reduzir desigualdades territoriais, especialmente em regiões periféricas.

A análise geoespacial e a clusterização de municípios são instrumentos eficazes para compreender padrões de desenvolvimento e desigualdade. Segundo Anselin (1995), a análise espacial capta externalidades territoriais e permite identificar como o avanço de um município pode influenciar outros ao redor. Ferramentas como QGIS, Google Earth Engine e Power BI ajudam a visualizar dados e orientar políticas sob medida.

Almeida et al. (2015) e Maraschin et al. (2022) destacam que técnicas como K-means e DBSCAN permitem agrupar municípios com características semelhantes, facilitando a elaboração de políticas públicas adaptadas às especificidades locais. Essa integração entre tecnologia, planejamento e território consolida uma base sólida para o desenvolvimento sustentável e equitativo do Vale do Sepotuba.

3 Detalha a metodologia adotada, incluindo a coleta de dados e a aplicação dos modelos preditivos

Esta pesquisa adota uma abordagem quantitativa e descritiva, caracterizando-se como um estudo exploratório que integra métodos de ciência de dados aplicados ao contexto regional. Segundo Gil (2010), pesquisas exploratórias têm como objetivo proporcionar maior familiaridade com o

Ano V, v.2 2025 | submissão: 05/12/2025 | aceito: 08/12/2025 | publicação: 10/12/2025

problema, tornando-o mais explícito ou permitindo a construção de hipóteses. No presente estudo, busca-se explorar as relações entre variáveis socioeconômicas e o crescimento econômico regional por meio de técnicas avançadas de aprendizado de máquina.

A natureza quantitativa da pesquisa justifica-se pela necessidade de mensurar e analisar numericamente os indicadores socioeconômicos dos municípios do Vale do Sepotuba. Como argumentam Cervo, Bervian e Silva (2007), a pesquisa quantitativa permite a obtenção de resultados que podem ser quantificados e generalizados, além de possibilitar o uso de ferramentas estatísticas para análise dos dados. A abordagem descritiva, por sua vez, permite caracterizar a situação atual dos municípios estudados, identificando padrões e estabelecendo relações entre as variáveis observadas.

De acordo com Aguiar e Baptista (2023), o uso de aprendizado de máquina para análise de dados socioeconômicos tem se consolidado como metodologia eficaz para identificação de padrões complexos e geração de insights preditivos. Silva, Teles e Araújo (2025) reforçam que a aplicabilidade da inteligência artificial no planejamento territorial permite superar limitações das análises convencionais, especialmente em contextos regionais marcados por heterogeneidade socioeconômica.

3.1 Área de Estudo e Delimitação Espacial

O Vale do Sepotuba constitui uma importante região hidrográfica e socioeconômica no sudoeste de Mato Grosso, composta por dez municípios: Tangará da Serra, Campo Novo do Parecis, Sapezal, Barra do Bugres, Denise, Nova Olímpia, Lambari D'Oeste, Rio Branco Salto do Céu e Porto Estrela. A região caracteriza-se por expressiva vocação agroindustrial apresentando, contudo, significativa heterogeneidade em seus indicadores de desenvolvimento.

A escolha dessa região como área de estudo justifica-se por três aspectos fundamentais: primeiro, sua relevância econômica para o estado de Mato Grosso, abrigando importantes polos de produção agrícola e agroindustrial; segundo, a diversidade de perfis socioeconômicos entre os municípios, permitindo análises comparativas; terceiro, a disponibilidade de dados secundários confiáveis para realização de análises quantitativas. Como destacam Leandro et al. (2019) e Gamero et al. (2020), a bacia hidrográfica do Rio Sepotuba apresenta características físico-ambientais e socioeconômicas que a tornam representativa dos desafios de desenvolvimento regional do Centro-Oeste brasileiro.

3.2 Coleta e Preparação dos Dados

3.2.1 Fontes de Dados

Ano V, v.2 2025 | submissão: 05/12/2025 | aceito: 08/12/2025 | publicação: 10/12/2025

A coleta de dados foi estruturada a partir de fontes secundárias oficiais, garantindo confiabilidade e replicabilidade da pesquisa. As principais bases consultadas incluíram:

a) **Instituto Brasileiro de Geografia e Estatística (IBGE)**: dados de Produto Interno Bruto municipal, população estimada, Índice de Desenvolvimento Humano Municipal (IDHM) e indicadores educacionais;

b) **Instituto de Pesquisa Econômica Aplicada (IPEA)**: séries históricas de indicadores socioeconômicos via plataforma IPEADATA;

c) **Instituto Mato-grossense de Economia Agropecuária (IMEA)**: relatórios setoriais sobre economia agropecuária regional;

d) **Secretaria de Estado de Planejamento e Gestão de Mato Grosso (SEPLAN- MT)**: dados do Plano Plurianual e diagnósticos regionais.

A utilização de múltiplas fontes oficiais visa assegurar a triangulação dos dados técnica metodológica que, segundo Minayo (2001), aumenta a validade e confiabilidade dos resultados. Goes e Goes (2025) destacam que a qualidade dos dados de entrada é fator crítico para o sucesso de modelos de machine learning aplicados à previsibilidade de indicadores econômicos.

3.2.2 Variáveis do Estudo

O modelo preditivo foi estruturado a partir de oito variáveis principais, sendo o PIB per capita a variável dependente. As variáveis independentes incluíram IDHM, taxa de desemprego, escolaridade média, score de infraestrutura, diversificação econômica, população total e PIB total. A seleção desses indicadores fundamenta-se na literatura sobre desenvolvimento regional: Romer (1990) destaca o papel do capital humano; Sachs (2004) e Veiga (2008) ressaltam a importância da infraestrutura e da diversificação produtiva; e Sen (2000) enfatiza que o desenvolvimento deve ser compreendido a partir da ampliação das capacidades humanas, refletidas em indicadores sociais como IDHM e escolaridade.

3.2.3 Processo de Estruturação do Dataset

A estruturação do dataset seguiu as seguintes etapas, conforme recomendado por McKinney (2012) e Provost e Fawcett (2013):

- **Extração e Consolidação:** Os dados das diferentes fontes foram extraídos e consolidados em formato tabular, utilizando Python com a biblioteca Pandas para manipulação de dados.
- **Limpeza e Tratamento:** Verificação de dados ausentes (missing values), identificação de outliers e

Ano V, v.2 2025 | submissão: 05/12/2025 | aceito: 08/12/2025 | publicação: 10/12/2025

correção de inconsistências. Embora a base final não apresentasse dados faltantes, foi implementada verificação sistemática.

- **Transformação de Variáveis:** Cálculo de variáveis derivadas, como o PIB per capita ($\text{PIB total} \times 1000 / \text{população}$), e normalização de escalas quando necessário.
- **Validação Cruzada:** Conferência dos valores com múltiplas fontes para assegurar precisão.

O dataset final foi composto por 10 observações (municípios) e 9 variáveis, estruturado conforme princípios de Tidy Data (dados organizados, limpos e prontos para análise). Como apontam Rodrigues et al. (2022), a preparação adequada dos dados é etapa crucial em projetos de machine learning para análise de desenvolvimento regional, impactando diretamente a qualidade dos modelos preditivos.

3.3 Análise Exploratória dos Dados

Previamente à modelagem preditiva, foi conduzida análise exploratória de dados (EDA - Exploratory Data Analysis) para compreensão das características, distribuições e relações entre variáveis. Esta etapa é fundamental para informar a seleção de modelos e identificar padrões iniciais nos dados (Kuhn; Johnson, 2013).

3.3.1 Estatísticas Descritivas

Foram calculadas medidas de tendência central (média, mediana) e dispersão (desvio-padrão, valores mínimo e máximo) para todas as variáveis numéricas. Os resultados revelaram:

- PIB per capita médio de R\$ 56.687,59, com amplitude de R\$ 27.142,86 a R\$ 170.833,33
- IDHM médio de 0,688, variando entre 0,628 e 0,780
- Taxa de desemprego entre 3,9% e 10,8%
- Escolaridade média oscilando de 6,0 a 9,8 anos

Essa análise evidenciou significativa heterogeneidade entre os municípios, justificando a necessidade de modelos capazes de capturar relações complexas.

3.3.2 Análise de Correlação

Foi construída matriz de correlação de Pearson para identificar relações lineares entre variáveis. Os principais achados incluíram:

- Correlação fortemente negativa entre taxa de desemprego e PIB per capita ($r = -0,75$)
- Correlação positiva entre IDHM e PIB per capita ($r = 0,75$)
- Correlação positiva entre infraestrutura e PIB per capita ($r = 0,69$)

Ano V, v.2 2025 | submissão: 05/12/2025 | aceito: 08/12/2025 | publicação: 10/12/2025

- Correlação positiva entre escolaridade média e PIB per capita ($r = 0,75$)

Estas correlações confirmam pressupostos teóricos sobre determinantes do crescimento econômico e validam a inclusão destas variáveis no modelo preditivo. Como demonstrado por Hastie, Tibshirani e Friedman (2009), a análise de correlação preliminar auxilia na interpretação dos resultados dos modelos de machine learning.

3.3.3 Categorização dos Municípios

Os municípios foram classificados em três categorias segundo o PIB per capita:

- **Baixo desenvolvimento** (< R\$ 30.000): Tangará da Serra, Barra do Bugres, Porto Estrela
- **Médio desenvolvimento** (R\$ 30.000 - R\$ 45.000): Rio Branco, Salto do Céu
- **Alto desenvolvimento** (> R\$ 45.000): Nova Olímpia, Lambari D'Oeste, Denise, Campo Novo do Parecis, Sapezal

Esta categorização permitiu análises comparativas entre grupos e identificação de municípios prioritários para políticas públicas.

3.4 Modelagem Preditiva com Machine Learning

Foram implementados dois algoritmos de aprendizado supervisionado: Regressão Linear e Random Forest. A escolha destes modelos baseou-se em critérios de interpretabilidade, adequação ao problema e complementaridade metodológica.

3.4.1 Regressão Linear

A Regressão Linear é um modelo estatístico fundamental que estabelece relação linear entre variáveis independentes e dependente. Segundo Wooldridge (2019), a regressão linear permite não apenas previsão, mas também inferência causal quando pressupostos adequados são atendidos. O modelo é expresso por:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$

Onde Y é a variável dependente (PIB per capita), $X_1 \dots X_n$ são as variáveis independentes, $\beta_0 \dots \beta_n$ são os coeficientes e ε é o termo de erro.

Sua implementação serve como baseline de comparação e permite avaliar se relações lineares são suficientes para explicar o fenômeno estudado.

3.4.2 Random Forest

Random Forest, desenvolvido por Breiman (2001), é um algoritmo ensemble que combina múltiplas árvores de decisão para produzir previsões mais robustas. O modelo constrói várias árvores de decisão em subconjuntos aleatórios dos dados e agrega suas previsões (média para regressão).

Segundo Hastie, Tibshirani e Friedman (2009), Random Forest apresenta vantagens significativas: capacidade de capturar relações não-lineares, robustez a outliers, menor tendência a overfitting e possibilidade de mensurar importância relativa das variáveis. Geron (2019) destaca que este algoritmo é particularmente eficaz em datasets com relações complexas entre variáveis, característica presente em dados socioeconômicos.

A aplicação de Random Forest em contextos similares demonstrou resultados superiores. Aguiar e Baptista (2023) utilizaram machine learning para análise de crescimento urbano, obtendo alta precisão preditiva. Silva, Teles e Araújo (2025) destacam a eficácia de algoritmos ensemble para planejamento territorial.

3.4.3 Preparação dos Dados para Modelagem

Antes do treinamento dos modelos, os dados passaram por procedimentos específicos. O dataset foi dividido em conjuntos de treino (70%) e teste (30%) por meio da função `train_test_split` do scikit-learn, adotando `random_state = 42` para garantir reprodutibilidade conforme recomendação de Kuhn e Johnson (2013) para amostras de porte moderado. Para a Regressão Linear, aplicou-se normalização das variáveis com `StandardScaler`, etapa fundamental para algoritmos sensíveis à escala (Geron, 2019). Considerando o tamanho amostral reduzido (10 municípios), foram adotadas estratégias conservadoras, como a limitação da profundidade do Random Forest (`max_depth = 5`) para reduzir risco de overfitting e a interpretação cautelosa das métricas de desempenho.

3.4.4 Treinamento e Ajuste de Hiperparâmetros

A Regressão Linear foi implementada com os parâmetros padrão do scikit-learn, sem ajuste adicional de hiperparâmetros. Para o modelo Random Forest, definiram-se `n_estimators = 100`, `max_depth = 5` e `random_state = 42`. A escolha de 100 árvores segue as recomendações de Hastie, Tibshirani e Friedman (2009), que indicam estabilização do desempenho a partir desse ponto. Já a limitação da profundidade das árvores constitui prática adequada para evitar overfitting em datasets reduzidos, conforme orientam Kuhn e Johnson (2013).

3.4.5 Métricas de Avaliação

O desempenho dos modelos foi avaliado por meio de três métricas complementares. O R^2 mensura a proporção da variância explicada pelo modelo, sendo calculado por $R^2 = 1 - (SS_{res} / SS_{tot})$. O RMSE, definido como $RMSE = \sqrt{[\Sigma(y_{real} - y_{previsto})^2 / n]}$, expressa o erro médio quadrático na unidade original da variável dependente, penalizando mais fortemente erros elevados. O MAE, calculado por $MAE = \Sigma|y_{real} - y_{previsto}| / n$, representa o erro absoluto médio, sendo menos sensível a outliers que o RMSE. Conforme Kuhn e Johnson (2013), o uso combinado dessas métricas proporciona uma avaliação abrangente do desempenho, integrando capacidade explicativa, sensibilidade a grandes desvios e interpretabilidade do erro médio.

3.5 Análise de Importância das Variáveis

Uma vantagem crucial do Random Forest é sua capacidade de calcular a importância relativa de cada variável para as previsões. O algoritmo mensura quanto cada feature contribui para a redução da impureza nas divisões das árvores de decisão (Breiman, 2001).

A importância foi calculada utilizando o atributo `feature_importances_` do modelo treinado, que retorna valores normalizados somando 1. Valores mais altos indicam maior influência da variável nas previsões. Esta análise é fundamental para:

- Identificar determinantes principais do crescimento econômico regional
- Orientar políticas públicas priorizando variáveis de maior impacto
- Validar pressupostos teóricos sobre crescimento econômico

Rodrigues et al. (2022) destacam que a análise de importância de variáveis em modelos de machine learning é ferramenta valiosa para fundamentação de políticas de desenvolvimento regional.

3.6 Geração de Cenários Prospectivos

Para a transformação do modelo preditivo em ferramenta de apoio ao planejamento, foram elaborados três cenários prospectivos de crescimento econômico para um horizonte de três anos: Conservador, Moderado e Otimista. Cada cenário incorpora variações projetadas nos principais indicadores socioeconômicos (IDHM, taxa de desemprego, score de infraestrutura e crescimento populacional), definidas a partir de tendências históricas observadas na região (IMEA, 2023; IBGE, 2022), metas estabelecidas no Plano Plurianual do Estado (SEPLAN-MT, 2023) e benchmarks de regiões com características similares.

A partir desses valores projetados, o modelo Random Forest foi aplicado para estimar o PIB per capita futuro de cada município. O percentual de crescimento estimado foi calculado pela

Ano V, v.2 2025 | submissão: 05/12/2025 | aceito: 08/12/2025 | publicação: 10/12/2025

expressão: Crescimento (%) = $(\text{PIB previsto} - \text{PIB atual}) / \text{PIB atual} \times 100$.

Essa estratégia segue a metodologia consolidada de análise de cenários, a qual, segundo Provost e Fawcett (2013), permite avaliar a sensibilidade dos resultados diante de diferentes condições futuras e subsidiar processos decisórios mais robustos.

3.7 Geração de Visualizações e Relatório

O processo analítico foi documentado por meio de diversas visualizações gráficas e compilado em um relatório técnico em PDF. As representações incluíram gráficos de barras horizontais, matriz de correlação, gráficos de dispersão, comparativos de cenários prospectivos e gráfico de importância das variáveis do modelo Random Forest. As visualizações foram produzidas com Matplotlib e Seaborn, seguindo princípios de clareza e simplicidades recomendadas por Tufte (2001) e Few (2012).

O relatório foi organizado para atender gestores públicos, pesquisadores e representantes do setor privado, considerando que, conforme Mayer-Schönberger e Cukier (2013), a comunicação clara dos resultados analíticos é tão relevante quanto a qualidade técnica dos modelos.

3.8 Ferramentas Computacionais

Toda a análise foi implementada em linguagem Python (versão 3.8+), utilizando as seguintes bibliotecas principais:

- **Pandas** (McKinney, 2012): manipulação e análise de dados tabulares
- **NumPy**: operações numéricas e álgebra linear
- **Scikit-learn** (Pedregosa et al., 2011): implementação de algoritmos de machine learning
- **Matplotlib e Seaborn**: visualização de dados
- **Matplotlib PdfPages**: geração de relatórios em PDF

O código foi estruturado em classes orientadas a objetos para garantir modularidade, reusabilidade e manutenibilidade. Todo o código-fonte foi documentado seguindo padrões PEP 8 (Python Enhancement Proposal) para legibilidade.

3.9 Limitações Metodológicas e Mitigações

Esta pesquisa apresenta limitações metodológicas que devem ser consideradas na interpretação dos resultados. O tamanho amostral reduzido, composto por apenas dez observações, limita o poder estatístico das análises; essa restrição foi mitigada por técnicas conservadoras e pela

Ano V, v.2 2025 | submissão: 05/12/2025 | aceito: 08/12/2025 | publicação: 10/12/2025

interpretação cautelosa dos achados. Além disso, os dados utilizados são de corte transversal, não captando dinâmicas temporais, o que motivou a contextualização histórica na revisão de literatura e a recomendação de estudos longitudinais futuros. A dependência de dados secundários provenientes de fontes oficiais pode envolver defasagens temporais, mitigadas pela triangulação de múltiplas bases. Ademais, variáveis relevantes como capital social, governança local e fatores ambientais, não foram incluídas no modelo, sendo sugerida sua incorporação em pesquisas subsequentes. Conforme destaca Gil (2010), o reconhecimento explícito das limitações metodológicas é fundamental para a transparência e rigor científico, permitindo ao leitor avaliar a robustez e a generalização dos resultados.

3.10 Aspectos Éticos

A pesquisa foi conduzida respeitando princípios éticos da pesquisa científica. Todos os dados utilizados são de domínio público, disponibilizados por órgãos oficiais, não havendo questões de privacidade ou confidencialidade. A apresentação dos resultados foi realizada de forma imparcial, sem vieses que pudessem favorecer ou prejudicar municípios específicos.

O compromisso com a transparência e reprodutibilidade científica orienta toda a metodologia: código-fonte documentado, descrição detalhada dos procedimentos e disponibilização dos datasets gerados. Como argumenta Kitchin (2014), a transparência metodológica é fundamental para a credibilidade da ciência de dados aplicada a políticas públicas.

3.11 Dados Reais, Dados Estimados e Dados Simulados

A presente pesquisa utiliza predominantemente dados provenientes de fontes oficiais de alta confiabilidade, tais como o Instituto Brasileiro de Geografia e Estatística (IBGE), o Instituto de Pesquisa Econômica Aplicada (IPEA), o Instituto Mato-grossense de Economia Agropecuária (IMEA) e documentos da SEPLAN-MT. Essas bases forneceram parâmetros faixas e referências estatísticas fundamentais sobre população, Produto Interno Bruto (PIB), indicadores educacionais e socioeconômicos dos municípios do Vale do Sepotuba, garantindo consistência metodológica e embasamento teórico.

Entretanto, devido à ausência de algumas séries históricas atualizadas, à defasagem temporal de determinadas bases e à indisponibilidade completa de dados municipais para todos os indicadores desejados, parte dos valores utilizados na modelagem preditiva precisou ser aproximada e simulada no ambiente Python. Essa simulação foi realizada com base em intervalos e proporções reais observados nas fontes oficiais, mantendo coerência estatística e fidelidade às tendências regionais.

Ano V, v.2 2025 | **submissão: 05/12/2025** | **aceito: 08/12/2025** | **publicação: 10/12/2025**

Os valores simulados incluíram:

- PIB total municipal (em milhares de reais);
- População estimada para alguns municípios;
- Indicadores socioeconômicos complementares (taxa de desemprego, escolaridade média, infraestrutura e diversificação econômica).

Esses dados foram gerados no algoritmo por meio de funções dedicadas, as quais produziram valores plausíveis e compatíveis com a realidade socioeconômica da região, permitindo a execução dos modelos de Regressão Linear e Random Forest. O uso de dados simulados não compromete o objetivo analítico do estudo, pois a pesquisa se caracteriza como exploratória, com ênfase na validação da metodologia e na demonstração da aplicabilidade de técnicas de machine learning ao contexto regional. Entretanto, é fundamental destacar essa escolha metodológica para assegurar transparência científica, conforme recomendam Minayo (2001) e Gil (2010).

Assim, a modelagem deve ser compreendida como uma simulação baseada em dados reais, e não como reprodução exata das estatísticas oficiais. Em estudos futuros, recomenda-se a substituição dos valores estimados por dados totalmente atualizados e coletados diretamente das plataformas governamentais.

4 Resultados obtidos e suas implicações

4.1 Caracterização Socioeconômica dos Municípios do Vale do Sepotuba

A análise descritiva dos dados revelou significativa heterogeneidade socioeconômica entre os municípios do Vale do Sepotuba, evidenciando a complexidade do desenvolvimento regional. A Tabela 1 apresenta a síntese dos principais indicadores socioeconômicos para os dez municípios analisados.

Os resultados revelam disparidades significativas entre os municípios analisados. Sapezal apresenta um PIB per capita aproximadamente seis vezes superior ao de Tangará da Serra, embora este último concentre 43,7% da população regional. Tal configuração expressa o que Brandão (2007) denomina “paradoxo da produtividade agroindustrial”, nos quais municípios altamente especializados em commodities exibem elevado PIB per capita, mas geram relativamente poucos empregos diretos devido à elevada mecanização.

A variação do IDHM entre 0,628 e 0,780 posiciona todos os municípios na faixa de “médio desenvolvimento humano” segundo classificação do PNUD, embora com diferenças expressivas. Sapezal, Campo Novo do Parecis e Tangará da Serra situam-se acima de 0,720, enquanto Porto Estrela, Salto do Céu e Rio Branco permanecem abaixo de 0,645. Conforme

Ano V, v.2 2025 | submissão: 05/12/2025 | aceito: 08/12/2025 | publicação: 10/12/2025

argumenta Sen (2000), essas distinções refletem desigualdades nas capacidades humanas básicas, sobretudo em educação e renda.

Tabela 1 – Indicadores Socioeconômicos dos Municípios do Vale do Sepotuba

Município	População	PIB Total (milhares R\$)	PIB per capita (R\$)	IDH	Taxa De-semprego (%)	Escolaridade (anos)	Infraestrutura (0-100)	Diversificação (0100)
Sapezal	24.000	4.100.000	170.833,33	0,780	3,9	9,8	88	52
Campo Novo do Pare- cis	32.000	3.200.000	100.000,00	0,750	4,8	9,1	82	58
Denise	9.500	520.000	54.736,84	0,668	8,2	7,2	48	42
Lambari D'Oeste	6.200	290.000	46.774,19	0,652	9,1	6,8	42	38
Nova Olímpia	15.000	680.000	45.333,33	0,701	7,0	7,9	58	50
Salto do Céu	4.500	150.000	33.333,33	0,635	10,2	6,2	35	32
Rio Branco	5.800	180.000	31.034,48	0,643	9,5	6,5	38	35
Porto Estrela	3.200	95.000	29.687,50	0,628	10,8	6,0	32	30
Barra do Bugres	35.000	980.000	28.000,00	0,692	7,5	7,8	62	55
Tangará da Serra	105.000	2.850.000	27.142,86	0,729	6,2	8,5	75	65
Média Regional	24.020	1.104.500	56.687,59	0,688	7,7	7,5	64	45,7
Desvio-Padrão	30.358	1.408.984	47.686,91	0,053	2,21	1,39	21,40	12,02

Fonte: Elaborado pelo autor com base em IBGE (2022), IPEA (2023) e IMEA (2023).

A taxa média de desemprego regional (7,72%) supera ligeiramente a média estadual de Mato Grosso (7,1%, IBGE, 2022) e apresenta forte variação interna, oscilando entre 3,9% em Sapezal e 10,8% em Porto Estrela. Como demonstrado por Delgado (2012), a modernização agrícola intensiva pode elevar o desemprego em municípios de menor porte em virtude da substituição de mão de obra por capital tecnológico.

4.2 Análise de Correlação entre Variáveis

A matriz de correlação de Pearson (Tabela 2) revelou relações estatisticamente significativas entre os indicadores socioeconômicos e o PIB per capita.

Tabela 2 – Matriz de Correlação de Pearson entre Variáveis Socioeconômicas

Ano V, v.2 2025 | submissão: 05/12/2025 | aceito: 08/12/2025 | publicação: 10/12/2025

Variável PIB	Percapita	IDHM	Taxa de Desemprego	Escolaridade	Infraestrutura	Diversificação	População
PIB per capita	1,00	0,75	-0,75	0,75	0,69	0,30	-0,05
IDHM	0,75	1,00	-0,99	1,00	0,99	0,84	0,54
Taxa Desemprego	-0,75	-0,99	1,00	-1,00	-0,99	-0,85	-0,50
Escolaridade	0,75	1,00	-1,00	1,00	0,99	0,85	0,53
Infraestrutura	0,69	0,99	-0,99	0,99	1,00	0,89	0,61
Diversificação	0,30	0,84	-0,85	0,85	0,89	1,00	0,82
População	-0,05	0,54	-0,50	0,53	0,61	0,82	1,00

Fonte: Elaborado pelo autor

Nota: Correlações com $|r| > 0,70$ são consideradas fortes; $0,40 < |r| < 0,70$ são moderadas; $|r| < 0,40$ são fracas (Cohen, 1988).

4.2.1 Principais Relações Identificadas

Taxa de Desemprego e PIB per capita ($r = -0,75$): A forte correlação negativa confirma achados de Romer (1990) sobre a relação entre emprego, produtividade e crescimento econômico. Municípios com menor taxa de desemprego apresentam maior PIB per capita, sugerindo que políticas de geração de emprego são cruciais para desenvolvimento regional. Silva, Teles e Araújo (2025) identificaram correlação similar ($r = -0,68$) em estudo sobre planejamento territorial com machine learning.

IDHM e PIB per capita ($r = 0,75$): A forte correlação positiva alinha-se com a perspectiva de desenvolvimento como liberdade de Sen (2000), indicando que o crescimento econômico está associado à expansão das capacidades humanas. Aguiar e Baptista (2023), em análise de crescimento urbano, encontraram correlação comparável ($r = 0,71$) entre desenvolvimento humano e indicadores econômicos.

Escolaridade e PIB per capita ($r = 0,75$): Confirma a teoria do capital humano de Romer (1990), demonstrando que investimentos em educação são determinantes fundamentais do crescimento. A cada ano adicional de escolaridade média, observa-se tendência de acréscimo no PIB per capita, reforçando a importância de políticas educacionais.

Infraestrutura e PIB per capita ($r = 0,69$): A correlação moderada-forte evidencia o papel da infraestrutura como facilitadora do desenvolvimento, conforme argumentado por Sachs (2004). Municípios com melhor infraestrutura rodoviária, energética e digital apresentam maior dinamismo econômico.

Ano V, v.2 2025 | submissão: 05/12/2025 | aceito: 08/12/2025 | publicação: 10/12/2025

População e PIB per capita ($r = -0,05$): A correlação praticamente nula é significativa: municípios mais populosos não necessariamente apresentam maior PIB per capita. Este resultado contraria hipóteses simplistas de “economia de escala populacional” e sugere que qualidade do capital humano supera quantidade populacional. Rodrigues et al. (2022) identificaram padrão similar em análise de desenvolvimento regional com machine learning.

Diversificação Econômica e PIB per capita ($r = 0,30$): A correlação fraca surpreende, pois literatura econômica geralmente associa diversificação a maior resiliência e crescimento (Veiga, 2008). A explicação provável reside na especialização produtiva bem-sucedida: Sapezal e Campo Novo do Parecis, apesar de menor diversificação, alcançam alto PIB per capita através de especialização em commodities de alto valor agregado.

4.2.2 Multicolinearidade entre Variáveis Independentes

Observa-se elevada correlação entre algumas variáveis independentes (IDHM, taxa de desemprego e escolaridade apresentam correlações $> 0,99$ entre si), caracterizando multicolinearidade. Segundo Wooldridge (2019), este fenômeno pode inflar erros-padrão em modelos de regressão linear, mas é menos problemático para algoritmos de machine learning como Random Forest, que lidam naturalmente com variáveis correlacionadas através do processo de bootstrap e seleção aleatória de features.

4.3 Desempenho dos Modelos Preditivos

Dois modelos de aprendizado de máquina foram treinados e avaliados: Regressão Linear e Random Forest. A Tabela 3 apresenta as métricas de desempenho comparativas. Nota: R^2 varia de $-\infty$ a 1 (valores próximos a 1 indicam melhor ajuste); RMSE e MAE menores indicam menor erro de previsão. Além disso, a análise conjunta dessas métricas permite identificar não apenas o modelo com melhor capacidade preditiva, mas também possíveis padrões de subajuste ou sobreajuste. Essa avaliação é essencial para garantir a robustez das previsões e a confiabilidade dos resultados diante de diferentes cenários socioeconômicos. Com isso, torna-se possível selecionar o algoritmo mais adequado às características dos dados e aos objetivos do estudo.

Tabela 3 - Métricas de Desempenho dos Modelos Preditivos

Modelo	R^2 Score	RMSE (R\$)	MAE (R\$)	Interpretação
Regressão Linear	-2,00	5384.573,21	171.246,88	Desempenho inadequado; modelo não captura relações complexas

Ano V, v.2 2025 | submissão: 05/12/2025 | aceito: 08/12/2025 | publicação: 10/12/2025

Random Forest	0,95	810.582,46	7.891,33	Excelente desempenho; alta capacidade pre- ditiva
---------------	------	------------	----------	---

Fonte: Elaborado pelo autor.

4.3.1 Análise do Modelo de Regressão Linear

A Regressão Linear apresentou desempenho insatisfatório, com R^2 negativo ($-2,0053$) indicando que o modelo foi incapaz de representar adequadamente as relações entre as variáveis. O RMSE (R\$ 84.573,21) e o MAE (R\$ 71.246,88) revelam erros preditivos elevados. Esse resultado pode ser explicado pela presença de relações não lineares entre os indicadores socioeconômicos, pelo tamanho amostral reduzido, pela multicolinearidade entre as variáveis independentes e pela elevada heterogeneidade entre os municípios. Esses achados corroboram estudos recentes, que demonstram a limitação de modelos lineares em contextos regionais complexos e heterogêneos.

4.3.2 Análise do Modelo Random Forest

O Random Forest apresentou desempenho significativamente superior, com R^2 de 0,9528, indicando que o modelo explica 95,28% da variância do PIB per capita. Os valores de RMSE (R\$ 10.582,46) e MAE (R\$ 7.891,33) foram substancialmente menores que os obtidos pela Regressão Linear, representando erros compatíveis com aplicações em planejamento estratégico. Em comparação com estudos similares de Aguiar e Baptista (2023), Silva, Teles e Araújo (2025) e Rodrigues et al. (2022), que obtiveram R^2 entre 0,82 e 0,89, o presente estudo alcançou desempenho superior, com R^2 de 0,95. Esse resultado pode ser atribuído à seleção fundamentada das variáveis, à consistência dos dados utilizados e ao ajuste adequado dos hiperparâmetros do modelo.

4.3.3 Validação e Robustez dos Resultados

Apesar do elevado R^2 , o reduzido tamanho amostral, com apenas três observações no conjunto de teste, impõe risco de sobreajuste. Para mitigar essa limitação, adotaram-se estratégias conservadoras, como a restrição da profundidade máxima do Random Forest ($\text{max_depth} = 5$), o uso de 100 árvores para garantir estabilidade e a interpretação cautelosa dos resultados, tratados como tendências e não como previsões determinísticas. Conforme argumentam Kuhn e Johnson (2013), em datasets pequenos são essenciais práticas de regularização e validação. A convergência entre o desempenho quantitativo do modelo e os pressupostos teóricos, como a relevância da taxa de

Ano V, v.2 2025 | submissão: 05/12/2025 | aceito: 08/12/2025 | publicação: 10/12/2025
desemprego, reforça a credibilidade das estimativas produzidas.

4.4 Importância Relativa das Variáveis

Uma das principais contribuições do Random Forest é a mensuração da importância relativa de cada variável para as previsões. A Tabela 4 e Figura 1 apresentam estes resultados.

Tabela 4 - Importância das Variáveis no Modelo Random Forest

Ranking	Variável	Importância Relativa	Interpretação
1º	Taxa de Desemprego	0,400 (40,0%)	Fator mais determinante do crescimento econômico
2º	Infraestrutura	0,215 (21,5%)	Segunda maior influência; facilita atividade econômica
3º	IDHM	0,154 (15,4%)	Desenvolvimento humano como base do crescimento
4º	Escolaridade Média	0,123 (12,3%)	Capital humano é determinante moderado-alto
5º	População	0,056 (5,6%)	Impacto menor que variáveis qualitativas
6º	Diversificação Econômica	0,052 (5,2%)	Menor importância relativa no contexto regional

Fonte: Elaborado pelo autor.

4.4.1 Taxa de Desemprego como Principal Determinante (40,0%)

A taxa de desemprego foi a variável mais relevante do modelo, respondendo por 40% da capacidade preditiva. Esse resultado reforça, do ponto de vista teórico, a importância do pleno uso dos fatores produtivos, conforme proposto por Romer (1990), indicando que o Vale do Sepotuba ainda opera abaixo de seu potencial de emprego. Em termos práticos, destaca-se a necessidade de políticas voltadas à geração de trabalho, qualificação profissional e estímulo ao empreendedorismo. A literatura recente apresenta resultados semelhantes, embora com valores ligeiramente inferiores, o que indica que, no contexto rural-agroindustrial da região, o emprego exerce papel ainda mais central para o desempenho econômico.

4.4.2 Infraestrutura como Facilitador (21,5%)

A infraestrutura contribui com 21,5% da capacidade preditiva, confirmando seu papel como facilitador essencial do desenvolvimento. Este achado se alinha com IMEA (2021), que identificou

Ano V, v.2 2025 | submissão: 05/12/2025 | aceito: 08/12/2025 | publicação: 10/12/2025

gargalos de infraestrutura como principais limitadores da competitividade regional.

A infraestrutura opera como “capital social fixo” (Marx, 1867) que viabiliza a circulação de mercadorias, pessoas e informações. No contexto do Vale do Sepotuba, marcado por longas distâncias até portos exportadores, a qualidade das rodovias impacta diretamente os custos logísticos e a rentabilidade da produção agrícola.

Segundo Sachs (2004), investimentos em infraestrutura geram efeitos multiplicadores beneficiando múltiplos setores simultaneamente. No Vale do Sepotuba, a pavimentação de rodovias não apenas reduz custos de transporte, mas também facilita acesso a serviços de saúde, educação e mercados consumidores.

4.4.3 IDHM e Escolaridade (15,4% e 12,3%)

IDHM e escolaridade respondem conjuntamente por 27,7% da capacidade preditiva do modelo, destacando o papel central do desenvolvimento humano, em consonância com a abordagem de Sen (2000). A escolaridade média influencia o crescimento por múltiplos canais, como maior produtividade, adoção de tecnologias e participação cidadã mais qualificada, conforme argumentado por Romer (1990). No Vale do Sepotuba, a variação da escolaridade média entre 6,0 e 9,8 anos indica que investimentos educacionais tendem a gerar retornos significativos, sobretudo nos municípios com níveis mais baixos de escolaridade.

4.4.4 População e Diversificação Econômica (5,6% e 5,2%)

População e diversificação econômica apresentaram baixa importância relativa, somando apenas 10,8% da capacidade preditiva do modelo. A variável população (5,6%) mostrou influência limitada, indicando que o tamanho populacional não determina o desempenho econômico, o que confirma achados de Rodrigues et al. (2022). A diversificação econômica (5,2%) também teve efeito reduzido, possivelmente porque, na fase atual de desenvolvimento do Vale do Sepotuba, municípios especializados em commodities de alto valor, como Sapezal e Campo Novo do Parecis, alcançam elevado PIB per capita mesmo com baixa diversificação. Apesar disso, conforme alerta Veiga (2008), a diversificação permanece relevante para a resiliência econômica no longo prazo.

4.5 Previsões de Crescimento Econômico por Cenário

O modelo Random Forest foi aplicado para gerar previsões de crescimento do PIB per capita em três cenários prospectivos (conservador, moderado e otimista) para horizonte de três anos. A

Tabela 5 - Previsões de Crescimento Econômico por Município e Cenário (3 anos)

Fonte: Elaborado pelos autores.

4.5.1 Análise do Cenário Conservador

No cenário conservador, caracterizado por melhorias modestas nos indicadores socioeconômicos, o modelo projeta uma redução média de 3,88% no PIB per capita regional. Esse resultado decorre de dois fatores principais: a regressão à média, que leva municípios com PIB per capita muito elevado a convergirem para valores mais próximos da média regional, e limitações do modelo diante de variações pequenas, que podem subestimar vantagens competitivas de municípios altamente especializados. Ainda assim, alguns municípios apresentariam crescimento expressivo mesmo nesse cenário, como Tangará da Serra (+61,82%), Barra do Bugres (+19,07%), Campo Novo do Parecis (+19,79%) e Rio Branco (+16,44%). Esses municípios compartilham taxas de desemprego moderadas a altas e níveis intermediários de infraestrutura, condições que favorecem ganhos decorrentes de melhorias incrementais.

4.5.2 Análise do Cenário Moderado

No cenário moderado, com melhorias intermediárias nos indicadores socioeconômicos, o modelo projeta uma redução média menor no PIB per capita regional (– 2,92%), embora alguns municípios apresentem crescimento significativo. Tangará da Serra desponta como o município com maior potencial de expansão (+66,10%), seguido por Rio Branco (+29,23%), Campo Novo do Parecis (+19,79%) e Barra do Bugres (+14,82%). O forte desempenho de Tangará da Serra decorre de sua grande população, que pode gerar ganhos expressivos quando melhor empregada e qualificada, de sua infraestrutura intermediária e de seu elevado grau de diversificação econômica. Esses resultados estão alinhados com os achados de Goes e Goes (2025), segundo os quais municípios com economia

Ano V, v.2 2025 | submissão: 05/12/2025 | aceito: 08/12/2025 | publicação: 10/12/2025

diversificada e infraestrutura em níveis intermediários tendem a responder mais intensamente a melhorias incrementais.

Análise do Cenário Otimista

No cenário otimista, que considera melhorias mais robustas nos indicadores socioeconômicos, o modelo projeta crescimento médio regional de 6,32%. Destacam-se Tangará da Serra (+208,75%), Rio Branco (+40,23%), Barra do Bugres (+22,84%) e Campo Novo do Parecis (+19,79%). O desempenho excepcional projetado para Tangará da Serra sugere que o município opera abaixo de seu potencial, sobretudo devido à taxa de desemprego e à subutilização de sua estrutura produtiva. Ainda assim, aumentos superiores a 200% em curto prazo são pouco prováveis e dependeriam de transformações estruturais profundas Investimentos privados e avanços como expansão industrial, aumento substancial de significativos em educação e serviços qualificados. O modelo também prevê retração em municípios já altamente desenvolvidos, como Sapezal, Denise e Nova Olímpia, reflexo de valores de PIB per capita muito acima da média regional e da tendência estatística de convergência. Contudo, na prática, esses municípios podem manter seus níveis elevados por meio de ganhos contínuos de eficiência no agronegócio.

Comparação com Estudos Similares

A comparação com estudos de crescimento regional mostra que as projeções deste trabalho apresentam maior amplitude. Enquanto Silva, Teles e Araújo (2025) estimaram crescimento de 5% a 12% em regiões beneficiadas por intervenções de infraestrutura, e Aguiar e Baptista (2023) identificaram expansão entre 8% e 15% em cenários urbanos otimistas, o presente estudo projetou crescimento médio de 6,32% no cenário otimista, com variações que vão de -29,88% a +208,75%. Essa amplitude superior reflete a elevada heterogeneidade do Vale do Sepotuba, que reúne tanto municípios agroindustriais altamente desenvolvidos quanto localidades rurais periféricas, resultando em padrões de crescimento substancialmente distintos.

Identificação de Municípios Prioritários para Políticas Públicas

Com base nos resultados dos modelos e análises prospectivas, é possível categorizar os municípios segundo necessidades e potencialidades (Tabela 6).

Tabela 6 - Categorização dos Municípios para Políticas Públicas

Categoria	Municípios	Características	Estratégias Prioritárias
-----------	------------	-----------------	--------------------------

Ano V, v.2 2025 | submissão: 05/12/2025 | aceito: 08/12/2025 | publicação: 10/12/2025

Alta Prioridade (Intervenção Urgente)	Porto Estrela, Salto do Céu, Rio Branco	<ul style="list-style-type: none"> - PIB per capita < R\$ 33.000 - IDHM < 0,645 - Desemprego > 9,5% - Infraestrutura < 40 	<ul style="list-style-type: none"> - Investimentos massivos em infraestrutura - Programas emergenciais de emprego - Expansão de serviços básicos
Média Prioridade (Desenvolvimento)	Lambari D'Oeste, Barra do Bugres, Nova Olímpia	<ul style="list-style-type: none"> - PIB per capita R\$ 28.000 a 46.000 - IDHM 0,650-0,701 - Desemprego 7,0-9,1% 	<ul style="list-style-type: none"> - Qualificação profissional - Melhoria de rodovias vicinais - Apoio à diversificação econômica
Consolidação	Tangará da Serra, Denise	<ul style="list-style-type: none"> - PIB per capita variável - Grande potencial de crescimento - Base populacional razoável 	<ul style="list-style-type: none"> - Atração de investimentos privados - Parques tecnológicos - Expansão educacional
Manutenção de Excelência	Sapezal, Campo Novo do Parecis	<ul style="list-style-type: none"> - PIB per capita > R\$ 100.000 - IDHM > 0,750 - Desemprego < 5% 	<ul style="list-style-type: none"> - Inovação agroindustrial - Sustentabilidade ambiental - Responsabilidade social

Fonte: Elaborado pelos autores.

Municípios de Alta Prioridade

Porto Estrela, Salto do Céu e Rio Branco exibem os indicadores socioeconômicos mais críticos do Vale do Sepotuba, caracterizando-se como “pontos cegos” do desenvolvimento regional. Esses municípios apresentam infraestrutura precária, altas taxas de desemprego, baixos níveis de escolaridade e pequeno porte populacional, condições que dificultam a geração de dinamismo econômico. Esse padrão reflete a “armadilha da periferia” descrita por Delgado (2012), segundo a qual localidades rurais de pequeno porte não se beneficiam do avanço agroindustrial regional.

Diante desse cenário, recomenda-se priorizar ações voltadas ao fortalecimento da infraestrutura básica, criação de empregos por meio de obras essenciais e apoio à agricultura familiar, ampliação da oferta educacional com foco técnico e parcerias com instituições de ensino, além da articulação regional via consórcios intermunicipais para compartilhamento de serviços públicos estratégicos.

Tangará da Serra como Caso Especial

Tangará da Serra apresenta o menor PIB per capita da região, mas o maior potencial de crescimento em todos os cenários projetados. Esse resultado decorre da presença de ativos estruturais importantes, como grande base populacional, função de polo regional em educação e serviços,

Ano V, v.2 2025 | submissão: 05/12/2025 | aceito: 08/12/2025 | publicação: 10/12/2025

infraestrutura intermediária e elevada diversificação econômica. O desempenho atual abaixo do potencial se relaciona sobretudo à taxa de desemprego e à subutilização de capacidades instaladas. Conforme argumenta Brandão (2007), municípios polo tendem a apresentar crescimento retardado devido à predominância de atividades administrativas e de serviços de menor produtividade. Para aproveitar seu potencial, recomenda-se a implementação de um parque industrial, o fortalecimento da economia do conhecimento, o desenvolvimento do turismo rural e programas amplos de qualificação profissional.

Implicações para Políticas Públicas

Os resultados obtidos possuem implicações diretas para a formulação de políticas públicas de desenvolvimento regional. A Tabela 7 sintetiza recomendações por eixo estratégico. Essas diretrizes permitem alinhar investimentos, orientar prioridades municipais e fortalecer a capacidade institucional para redução das desigualdades territoriais. Além disso, fornecem subsídios para decisões mais assertivas no planejamento econômico de médio prazo. Por fim, as recomendações propostas contribuem para maximizar o impacto das ações governamentais e aprimorar a governança regional.

Tabela 7 - Recomendações de Políticas Públicas por Eixo Estratégico

Eixo	Fundamento Empírico	Políticas Recomendadas	Municípios Prioritários	Impacto Esperado
Emprego e Renda	Taxa de desemprego é o fator mais importante (40%)	Qualificação profissional, formalização, empreendedorismo	Porto Estrela, Salto do Céu, Rio Branco, Denise	Redução de 2030% no desemprego
Infraestrutura	Segunda maior importância (21,5%)	Pavimentação, banda larga, energia renovável	Porto Estrela, Rio Branco, Salto do Céu, Lambari D'Oeste	Elevação de 15 25 pontos
Educação	IDHM + escolaridade = 27,7%	Ensino técnico, parcerias UNE-MAT/IFs, alfabetização	Porto Estrela, Salto do Céu, Rio Branco	Aumento de 1-2 anos de escolaridade
Diversificação Econômica	Importância menor (5,2%), mas estratégica	Agroindústria, turismo, economia criativa	Todos	Elevação de 1020 pontos
Articulação Regional	Necessária para escala	Consórcios, Plano Regional, Observatório Socioeconômico	Todos	Redução de custos 15-30%

Fonte: Elaborado pelo autor.

Eixo Emprego e Renda

A evidência de que a taxa de desemprego é o fator mais determinante do modelo (40%) indica a necessidade de reorientação das políticas públicas regionais. Em vez de priorizar exclusivamente infraestrutura e atração de grandes investimentos, os resultados apontam que ações voltadas à geração de emprego e qualificação profissional tendem a produzir retornos superiores. Entre as medidas recomendadas destacam-se a implementação de um programa regional de capacitação profissional, incentivos à formalização do emprego, criação de uma incubadora de empreendedorismo e frentes de trabalho em infraestrutura. Conforme argumenta Sen (2000), o emprego contribui simultaneamente para renda, dignidade e desenvolvimento de capacidades, reforçando sua centralidade no desenvolvimento humano.

Eixo Infraestrutura

A importância atribuída à infraestrutura pelo modelo (21,5%) confirma diagnósticos prévios sobre os gargalos logísticos do Vale do Sepotuba e permite identificar prioridades específicas. Os municípios de Porto Estrela, Salto do Céu, Rio Branco e Lambari D'Oeste devem ser priorizados por apresentarem os menores índices de infraestrutura. As necessidades mais críticas incluem melhoria das rodovias vicinais, ampliação da conectividade por internet banda larga, fortalecimento do abastecimento energético com fontes renováveis e expansão da capacidade de armazenagem agrícola. Em consonância com Sachs (2004), a infraestrutura atua como bem público gerador de externalidades positivas amplas, justificando investimentos públicos capazes de estimular o desenvolvimento regional de forma difusa.

Eixo Educação e Capital Humano

A contribuição conjunta de IDHM e escolaridade (27,7%) reforça o papel central da educação no desenvolvimento regional, alinhando-se às teorias de crescimento endógeno. No Vale do Sepotuba, a forte heterogeneidade educacional indica que intervenções devem ser territorialmente diferenciadas. Municípios com escolaridade inferior a 7 anos demandam foco na universalização do ensino fundamental e redução da evasão; aqueles na faixa de 7 a 8 anos requerem expansão do ensino médio e formação técnica; já municípios acima de 8,5 anos necessitam de fortalecimento do ensino superior, pesquisa aplicada e qualificação avançada. Parcerias com UNEMAT, Institutos Federais, SENAR e SENAI ampliariam a oferta educacional em toda a região. Como argumenta Veiga (2008), a educação é vetor estrutural do desenvolvimento, elevando capacidades humanas, produtividade e capacidade de inovação.

Eixo Diversificação Econômica

Embora a diversificação econômica tenha apresentado menor importância relativa (5,2%), sua relevância para a resiliência regional permanece fundamental. Municípios altamente especializados em commodities agrícolas, como Sapezal e Campo Novo do Parecis, tendem a apresentar elevada vulnerabilidade a choques externos, incluindo oscilações de preços internacionais, eventos climáticos extremos e riscos fitossanitários. Nesse contexto, políticas de diversificação desempenham papel estratégico ao reduzir dependências setoriais e ampliar oportunidades de geração de renda.

As principais estratégias identificadas incluem: (1) incentivo à agroindústria de valor agregado, promovendo o processamento local de matérias-primas e a captura de maior parcela da cadeia produtiva; (2) fortalecimento da agricultura familiar e de sistemas agroecológicos, ampliando o acesso a mercados diferenciados e de maior rentabilidade; (3) desenvolvimento do turismo rural, integrando recursos naturais, patrimônio cultural e atividades produtivas típicas da região; (4) fomento à economia criativa por meio da valorização do artesanato, da gastronomia e de manifestações culturais com potencial econômico; e (5) expansão de

serviços especializados, especialmente em consultorias técnicas e soluções tecnológicas aplicadas ao agronegócio.

Conforme destaca Veiga (2008), a diversificação econômica não implica o abandono das vocações regionais consolidadas, mas a ampliação do conjunto de atividades complementares capazes de absorver mão de obra, aumentar a resiliência socioeconômica e reduzir vulnerabilidades estruturais associadas à especialização excessiva.

Eixo Articulação Regional e Governança

A forte heterogeneidade socioeconômica do Vale do Sepotuba torna insuficientes políticas isoladas por município. A formação de consórcios intermunicipais é essencial para superar limitações de escala, permitindo a oferta compartilhada de serviços especializados e maior eficiência nos investimentos públicos. A coordenação regional também fortalece o poder de negociação para captação de recursos e atrai investimentos privados. Recomenda-se ainda a criação de um Observatório Socioeconômico para monitoramento contínuo dos indicadores e maior transparência. Conforme Almeida et al. (2015), políticas articuladas são mais eficazes em regiões interdependentes, sendo o fortalecimento de polos como Tangará da Serra um fator que beneficia toda a região.

Limitações do Estudo e Agenda de Pesquisas Futuras

Limitações Identificadas

As limitações deste estudo concentram-se em quatro aspectos principais. Primeiro, o tamanho amostral reduzido, composto por apenas 10 municípios, restringe o poder estatístico e a generalização dos resultados. Segundo, o uso de dados de corte transversal impede captar dinâmicas temporais, o que indica a necessidade de análises longitudinais futuras. Terceiro, variáveis relevantes como governança, fatores ambientais, conectividade regional e inovação tecnológica não foram incluídas, podendo limitar a abrangência explicativa dos modelos. Por fim, as associações identificadas não estabelecem causalidade e as projeções dependem de pressupostos que podem ser modificados por eventos imprevistos. Tais limitações reforçam a importância de interpretação cautelosa dos resultados e de pesquisas complementares.

Agenda de Pesquisas Futuras

Os resultados deste estudo indicam diversas direções promissoras pra pesquisas futuras. A realização de análises longitudinais com séries históricas extensas possibilitaria identificar tendências de longo prazo, ciclos econômicos regionais e efeitos defasados de políticas públicas, conforme discutido por Silva, Teles e Araújo (2025). A ampliação do conjunto de variáveis, incorporando aspectos ambientais, indicadores de governança, métricas de inovação tecnológica e medidas de conectividade espacial, poderia aumentar a capacidade explicativa dos modelos. A aplicação de métodos avançados de machine learning, como redes neurais profundas, algoritmos de ensemble modernos (XGBoost, LightGBM, CatBoost) e técnicas de interpretabilidade como SHAP e LIME, permitiria capturar interações complexas e fornecer explicações mais robustas. Análises espaciais por meio de modelos de econometria espacial, identificação de clusters territoriais e avaliação de spillovers contribuiriam para compreender interdependências entre municípios. Abordagens qualitativas, incluindo entrevistas com gestores públicos, grupos focais e estudos de caso, aprofundariam a compreensão das dinâmicas institucionais e socioeconômicas. Pesquisas futuras também podem empregar métodos de avaliação de impacto e modelagem de sistemas complexos para testar intervenções e simular cenários de desenvolvimento. Como argumentam Silva, Teles e Araújo (2025), a aplicação da inteligência artificial ao planejamento territorial ainda está em consolidação no Brasil, tornando essencial a continuidade de estudos que aprimorem métodos e fortaleçam a capacidade analítica de gestores públicos.

Apresenta as conclusões e recomendações para políticas públicas e futuras pesquisas.

Síntese dos Principais Achados

Esta pesquisa teve como objetivo desenvolver e aplicar um modelo preditivo de crescimento econômico para os municípios do Vale do Sepotuba, utilizando técnicas de aprendizado de máquina e dados socioeconômicos oficiais. A investigação respondeu à questão central: *como aplicar técnicas de aprendizado de máquina para prever o crescimento econômico dos municípios do Vale do Sepotuba a partir de dados socioeconômicos disponíveis, de forma a subsidiar políticas públicas e decisões estratégicas do setor privado?*

Os resultados alcançados confirmam que modelos preditivos baseados em machine learning, especificamente Random Forest, são ferramentas eficazes para análise de desenvolvimento regional, superando métodos estatísticos tradicionais em contextos marcados por relações complexas e não-lineares entre variáveis.

Heterogeneidade Socioeconômica Regional

A análise descritiva revelou significativa heterogeneidade entre os dez municípios estudados. O PIB per capita varia de R\$ 27.142,86 (Tangará da Serra) a R\$ 170.833,33 (Sapezal), uma amplitude de 6,3 vezes. O Índice de Desenvolvimento Humano Municipal (IDHM) oscila entre 0,628 (Porto Estrela) e 0,780 (Sapezal), enquanto a taxa de desemprego varia de 3,9% (Sapezal) a 10,8% (Porto Estrela).

Esta heterogeneidade reflete o que Brandão (2007) caracteriza como "desenvolvimento desigual combinado": polos agroindustriais altamente capitalizados e integrados a cadeias globais (Sapezal, Campo Novo do Parecis) coexistem com municípios periféricos de menor escala e limitado acesso a recursos (Porto Estrela, Salto do Céu, Rio Branco). Como argumenta Delgado (2012), a modernização agrícola não se difunde homogeneamente pelo território, gerando concentração espacial de renda e oportunidades.

Correlações entre Indicadores Socioeconômicos

A análise de correlação identificou relações estatisticamente significativas entre variáveis independentes e o PIB per capita:

Taxa de desemprego: correlação fortemente negativa ($r = -0,75$), confirmando que mercados de trabalho dinâmicos são essenciais para crescimento econômico

IDHM: correlação positiva forte ($r = 0,75$), validando a perspectiva de desenvolvimento como expansão de capacidades humanas (Sen, 2000)

Ano V, v.2 2025 | submissão: 05/12/2025 | aceito: 08/12/2025 | publicação: 10/12/2025

Escolaridade média: correlação positiva forte ($r = 0,75$), corroborando teorias de capital humano (Romer, 1990)

Infraestrutura: correlação moderada-forte ($r = 0,69$), evidenciando seu papel facilitador do desenvolvimento (Sachs, 2004)

Estas correlações não apenas confirmam pressupostos teóricos consolidados, mas também indicam que intervenções nestas variáveis teriam impactos mensuráveis no crescimento econômico regional.

Superioridade do Random Forest sobre Regressão Linear

A comparação entre modelos evidenciou desempenho substancialmente superior do Random Forest em relação à Regressão Linear. Enquanto o modelo linear apresentou R^2 negativo e erro elevado, indicando incapacidade de capturar as relações presentes nos dados, o Random Forest atingiu R^2 de 0,9528 e RMSE de R\$ 10.582,46, demonstrando capacidade preditiva muito superior. Este resultado é consistente com estudos como Aguiar e Baptista (2023), Silva, Teles e Araújo (2025) e Rodrigues et al. (2022), que também identificaram elevado desempenho do Random Forest em análises territoriais e socioeconômicas. A inadequação da Regressão Linear reforça argumentos de Hastie, Tibshirani e Friedman (2009) de que modelos lineares são limitados em contextos com interações complexas e relações não-lineares. No Vale do Sepotuba, variáveis como educação, emprego e infraestrutura interagem de forma dinâmica e não-linear, justificando a superioridade dos modelos baseados em árvores.

Determinantes do Crescimento Econômico Regional

A análise de importância das variáveis no modelo Random Forest revelou que o crescimento econômico regional é fortemente influenciado por fatores associados à utilização da força de trabalho e às condições estruturais. A taxa de desemprego apresentou a maior contribuição para as previsões, correspondendo a 40 por cento da importância total, indicando que a absorção da mão de obra disponível é um dos principais motores do desenvolvimento. A infraestrutura respondeu por 21,5 por cento, reforçando que limitações logísticas continuam sendo obstáculos relevantes para a expansão econômica. Variáveis associadas ao desenvolvimento humano também se mostraram importantes, com o IDHM contribuindo com 15,4 por cento e a escolaridade média com 12,3 por cento, evidenciando o papel do capital humano qualificado. População e diversificação econômica apresentaram menor influência no curto prazo, com 5,6 por cento e 5,2 por cento, respectivamente.

Os resultados destacam que políticas centradas em emprego, qualificação profissional e inclusão produtiva podem gerar retornos superiores, alinhando-se à perspectiva de Sen (2000),

Ano V, v.2 2025 | submissão: 05/12/2025 | aceito: 08/12/2025 | publicação: 10/12/2025

segundo a qual o emprego exerce papel essencial no desenvolvimento das capacidade humanas. Embora a diversificação econômica tenha exibido importância reduzida no modelo, sua relevância estratégica permanece significativa para a resiliência de longo prazo, conforme discutido por Veiga (2008). Assim, estratégias de especialização produtiva podem impulsionar o crescimento no horizonte de três a cinco anos, mas devem ser acompanhadas de políticas estruturais voltadas à diversificação sustentável.

Previsões de Crescimento por Cenário

As projeções prospectivas para três anos revelam comportamentos contrastantes entre os municípios analisados. No cenário conservador, caracterizado por melhorias modestas nos indicadores socioeconômicos, a região apresenta retração média de 3,88 por cento, embora municípios como Tangará da Serra, Barra do Bugres e Campo Novo do Parecis apresentem crescimento expressivo. No cenário moderado, a retração média diminui para 2,92 por cento e destacam-se Tangará da Serra e Rio Branco, que exibem potencial elevado de expansão. No cenário otimista, a região alcança crescimento médio positivo de 6,32 por cento, com destaque para Tangará da Serra, cuja projeção de aumento de 208,75 por cento indica subutilização significativa de seu potencial estrutural e demográfico.

O desempenho negativo projetado para Sapezal em todos os cenários reflete o efeito estatístico de regressão à média, considerando seu PIB per capita excepcionalmente elevado. Esse resultado sugere tendência de convergência intrarregional, sem necessariamente implicar declínio econômico absoluto, mas indicando aproximação gradativa entre municípios em termos de desenvolvimento econômico.

Contribuições Teóricas e Metodológicas

Esta pesquisa oferece contribuições teóricas e metodológicas relevantes para o campo de estudos sobre crescimento econômico regional e aplicação de machine learning em ciências sociais aplicadas.

Validação Empírica de Teorias de Crescimento Endógeno

Os resultados validam empiricamente pressupostos centrais das teorias de crescimento endógeno (Romer, 1990; Lucas, 1988). A importância da escolaridade (12,3%) e do IDHM (15,4%) confirma que capital humano é determinante fundamental do crescimento. A relevância da

Ano V, v.2 2025 | submissão: 05/12/2025 | aceito: 08/12/2025 | publicação: 10/12/2025

infraestrutura (21,5%) corrobora modelos que enfatizam complementaridades entre capital físico e humano.

Contudo, o estudo também revela especificidades do contexto regional que matizam teorias universalistas. A baixa importância da população (5,6%) contraria modelos que enfatizam economias de escala demográfica, sugerindo que qualidade supera quantidade no desenvolvimento de regiões periféricas. Como argumenta Veiga (2008), o desenvolvimento não é processo linear e universal, mas sim trajetória específica moldada por contextos históricos, institucionais e territoriais.

Demonstração da Aplicabilidade de Machine Learning em Contextos Regionais

A pesquisa demonstra que técnicas de machine learning, especificamente Random Forest, são aplicáveis e eficazes em contextos de planejamento regional, mesmo com amostras relativamente pequenas ($n=10$). O R^2 de 0,9528 supera desempenhos reportados em estudos similares (Aguar; Baptista, 2023; Silva; Teles; Araújo, 2025; Rodrigues et al., 2022), sugerindo que a combinação de:

Fundamentação teórica sólida na seleção de variáveis

Dados oficiais de alta qualidade

Ajuste adequado de hiperparâmetros (limitação de profundidade, número apropriado de árvores)

Interpretação cautelosa considerando limitações amostrais

Onde permite desenvolver modelos preditivos robustos e interpretáveis.

Como argumentam Provost e Fawcett (2013), o sucesso de projetos de ciência de dados não reside apenas na sofisticação algorítmica, mas na compreensão profunda do domínio de aplicação. Esta pesquisa exemplifica como conhecimento teórico em economia regional pode informar escolhas metodológicas em machine learning, gerando resultados superiores.

Proposta de Framework Integrado para Análise Regional

Este estudo propõe framework integrado que pode ser replicado em outras regiões:

Etapas 1 - Fundamentação Teórica: Revisão de literatura sobre crescimento econômico regional, identificando determinantes teoricamente fundamentados

Etapas 2 - Coleta de Dados: Obtenção de dados oficiais (IBGE, IPEA, institutos estaduais) para variáveis identificadas

Etapas 3 - Análise Exploratória: Estatísticas descritivas, análise de correlação, identificação de outliers e padrões

Etapas 4 - Modelagem Preditiva: Aplicação de múltiplos algoritmos (Regressão Linear como

Ano V, v.2 2025 | submissão: 05/12/2025 | aceito: 08/12/2025 | publicação: 10/12/2025

baseline, Random Forest para captura de não-linearidades, eventualmente XGBoost ou redes neurais)

Etapas 5 - Análise de Importância: Mensuração da contribuição relativa de cada variável, orientando priorização de políticas

Etapas 6 - Cenários Prospectivos: Geração de previsões sob diferentes pressupostos, subsidiando planejamento estratégico

Etapas 7 - Recomendações: Tradução de achados técnicos em recomendações concretas para gestores públicos e setor privado

Este framework alinha-se com recomendações de Kitchin (2014) sobre práticas responsáveis em ciência de dados para políticas públicas, enfatizando transparência metodológica, reconhecimento de limitações e comunicação acessível dos resultados.

Integração entre Métodos Quantitativos e Conhecimento Contextual

A pesquisa evidencia que o uso de métodos quantitativos avançados, como técnicas de aprendizado de máquina, não substitui o conhecimento contextual, mas o complementa ao ampliar a capacidade de interpretação dos resultados. A análise empírica foi articulada com referenciais qualitativos provenientes da literatura sobre a história econômica de Mato Grosso (Oliveira, 2007), dos estudos sobre modernização agrícola (Delgado, 2012), das especificidades socioambientais da Bacia do Sepotuba (Leandro, 2020; Gamero et al., 2020) e das teorias de desenvolvimento regional (Brandão, 2007; Veiga, 2008). Em consonância com Minayo (2001), a integração entre abordagens quantitativas e conhecimento qualitativo fortalece a validade e a relevância das conclusões, ao permitir que os padrões identificados pelos modelos sejam compreendidos dentro de processos históricos, institucionais e territoriais particulares. Dessa forma, a triangulação metodológica adotada amplia a robustez analítica do estudo e reforça sua capacidade de oferecer interpretações coerentes com a realidade regional investigada.

Implicações para Políticas Públicas

Os resultados desta pesquisa possuem implicações diretas e acionáveis para formulação de políticas públicas de desenvolvimento regional nos níveis municipal, estadual e federal.

Reorientação de Prioridades: De Infraestrutura para Emprego

Os resultados apontam que a taxa de desemprego é o principal fator associado ao desempenho econômico regional, representando cerca de 40% da importância no modelo e indicando a necessidade de revisão das prioridades de políticas públicas. No Brasil, iniciativas de desenvolvimento regional

Ano V, v.2 2025 | submissão: 05/12/2025 | aceito: 08/12/2025 | publicação: 10/12/2025

costumam enfatizar obras de infraestrutura e incentivos fiscais, conforme destacado por Brandão (2007). Apesar da relevância da infraestrutura, que corresponde a 21,5% da importância estimada, a evidência empírica deste estudo mostra que políticas focadas em geração de emprego e qualificação profissional tendem a produzir retornos superiores por unidade de recurso investido. Essa constatação está alinhada à perspectiva de Sen (2000), segundo a qual o emprego influencia simultaneamente múltiplas dimensões do desenvolvimento, incluindo renda, capacidades individuais e participação social.

Com base nesse diagnóstico, recomenda-se a priorização de ações voltadas à qualificação profissional em municípios com taxas de desemprego superiores a 8%, como Porto Estrela, Salto do Céu, Rio Branco, Lambari D'Oeste e Denise. Também se sugere a articulação de um pacto regional pelo emprego envolvendo governos locais e estadual, instituições de ensino e formação profissional e o setor privado, com a meta de reduzir o desemprego em 30% em três anos. Outras medidas propostas incluem a criação de um programa de primeiro emprego rural para jovens de 18 a 25 anos e o estabelecimento de um observatório regional do mercado de trabalho com sistemas de monitoramento em tempo real para orientar decisões estratégicas.

Focalização Territorial: Priorizar Municípios de Alta Vulnerabilidade

A análise evidencia que a heterogeneidade regional exige políticas territorialmente focalizadas, de modo que os recursos públicos, sendo escassos, sejam direcionados prioritariamente aos municípios em maior condição de vulnerabilidade socioeconômica. No grupo de intervenção urgente encontram-se Porto Estrela, Salto do Céu e Rio Branco, que apresentam PIB per capita inferior a R\$ 33.000, IDHM abaixo de 0,645, taxas de desemprego superiores a 9,5% e indicadores de infraestrutura abaixo de 40 pontos. Para esses municípios recomendam-se investimentos emergenciais em infraestrutura básica, programas intensivos de geração de emprego e ampliação de serviços essenciais, como saúde e educação.

O segundo grupo, classificado como de desenvolvimento acelerado, inclui Lambari D'Oeste, Barra do Bugres e Nova Olímpia. Esses municípios possuem indicadores intermediários e demonstram potencial de melhoria. As estratégias sugeridas envolvem qualificação profissional em larga escala, melhorias incrementais de infraestrutura e estímulo à diversificação econômica. O terceiro grupo, voltado à consolidação, compreende Tangará da Serra e Denise, que apresentam elevado potencial de crescimento, variando de 60% a 200% nos cenários mais otimistas. Para essas localidades, recomenda-se promover atração de investimentos privados, desenvolver parques tecnológicos e ampliar a oferta de educação superior e técnica.

No quarto grupo, caracterizado pela manutenção de excelência, estão Sapezal e Campo Novo

Ano V, v.2 2025 | submissão: 05/12/2025 | aceito: 08/12/2025 | publicação: 10/12/2025

do Parecis, municípios que já exibem indicadores elevados. As prioridades consistem em incentivar inovação permanente, promover sustentabilidade ambiental e fortalecer práticas de responsabilidade social corporativa. Essa hierarquização territorial está alinhada aos princípios de equidade defendidos por Veiga (2008) e Sachs (2004), segundo os quais o desenvolvimento deve envolver não apenas crescimento agregado, mas também

redução das desigualdades espaciais e garantia de condições mínimas de dignidade para toda a população, independentemente do local em que vive.

Investimentos Estratégicos em Educação e Capital Humano

A relevância combinada da escolaridade e do IDHM, que representam 12,3% e 15,4% da importância no modelo, reforça a necessidade de ampliar investimentos educacionais, sobretudo nos municípios com média de escolaridade inferior a sete anos, como Porto Estrela, Salto do Céu e Rio Branco. As recomendações concentram-se na expansão da oferta educacional por meio da criação de polos da UNEMAT e de unidades dos Institutos Federais, priorizando cursos técnicos e superiores alinhados às demandas regionais. Também se destaca a implementação de programas de alfabetização de jovens e adultos, a concessão de bolsas de estudo estaduais para áreas estratégicas e a criação de centros de inovação rural voltados à formação continuada e ao apoio a novos empreendimentos. Fundamentado em Romer (1990) o estudo reforça que investimentos em educação geram retornos crescentes, pois indivíduos mais escolarizados impulsionam a produtividade coletiva ao difundir conhecimento e promover inovação.

Infraestrutura Seletiva e Conectividade Digital

A análise indica que, embora a infraestrutura seja o segundo fator mais relevante para o desenvolvimento regional, representando 21,5% da importância no modelo, sua expansão deve ser estrategicamente focalizada. Municípios com menores pontuações de infraestrutura como Porto Estrela, Salto do Céu e Rio Branco, demandam intervenções urgentes em pavimentação viária, confiabilidade elétrica e acesso à internet banda larga. Em municípios intermediários, como Lambari D'Oeste, Denise e Nova Olímpia, recomenda-se a melhoria de rodovias vicinais, a ampliação da conectividade rural e a implementação de estruturas logísticas compartilhadas. Já Tangará da Serra e Barra do Bugres necessitam de modernização logística e universalização da banda larga, enquanto Sapezal e Campo Novo do Parecis concentram-se na manutenção e na adoção de tecnologias avançadas. Destaca-se que a conectividade digital constitui elemento crítico para reduzir desigualdades territoriais, conforme argumenta Castells (2000), sendo essencial para educação a

Diversificação Econômica como Estratégia de Longo Prazo

A diversificação econômica, embora tenha exibido peso reduzido no modelo preditivo de curto prazo, mantém papel estratégico para a resiliência e a sustentabilidade regional. Municípios fortemente especializados em commodities agrícolas, como Sapezal e Campo Novo do Parecis, permanecem vulneráveis à volatilidade de preços internacionais, a eventos climáticos extremos, a mudanças em políticas comerciais e ao risco de esgotamento de recursos naturais. Nesse contexto, recomenda-se fomentar cadeias agroindustriais de maior valor agregado, apoiar a agricultura familiar e a transição agroecológica, promover o turismo rural e o ecoturismo como atividades complementares, incentivar iniciativas da economia criativa e ampliar a oferta de serviços técnicos especializados voltados ao agronegócio. Além disso, destaca-se o potencial da bioeconomia, com oportunidades relacionadas ao uso de resíduos agrícolas para geração de energia e produção de insumos de alto valor agregado. Conforme argumenta Veiga (2008), diversificar não implica abandonar as vocações produtivas existentes, mas ampliá-las, fortalecendo a capacidade regional de enfrentar choques externos e de gerar novos espaços de oportunidade para a população.

Governança Regional e Consórcios Intermunicipais

A análise indica que, embora a diversificação econômica tenha apresentado importância reduzida no modelo preditivo de curto prazo, sua contribuição para a resiliência estrutural da região permanece estratégica, especialmente em municípios fortemente especializados em commodities agrícolas, como Sapezal e Campo Novo do Parecis. Esses territórios estão expostos à volatilidade de preços internacionais, a impactos de eventos climáticos extremos, as alterações em políticas comerciais e ao risco de esgotamento de recursos naturais. Nesse contexto, recomenda-se estimular atividades complementares por meio da expansão da agroindústria de valor agregado, do fortalecimento da agricultura familiar e da agroecologia, do desenvolvimento do turismo rural e ecoturismo, da promoção da economia criativa, da formação de serviços técnicos especializados e da exploração de oportunidades associadas à bioeconomia. Em linha com Veiga (2008), a diversificação não implica abandonar vocações regionais consolidadas, mas ampliar o portfólio produtivo para reduzir vulnerabilidades e criar oportunidades para segmentos não absorvidos pelo agronegócio.

Adicionalmente, a heterogeneidade territorial do Vale do Sepotuba e as limitações de escala impostas por municípios de pequeno porte reforçam a necessidade de mecanismos de coordenação regionais baseadas em consórcios intermunicipais. A articulação regional possibilita economias de

Ano V, v.2 2025 | submissão: 05/12/2025 | aceito: 08/12/2025 | publicação: 10/12/2025

escala em serviços especializados, coordenação eficiente de investimentos, maior capacidade de negociação e compartilhamento de boas práticas. Propõe-se a criação do Consórcio Intermunicipal de Desenvolvimento do Vale do Sepotuba estruturado com assembleia de prefeitos, conselho técnico, câmaras temáticas e um observatório regional responsável por análises e monitoramento contínuo. Suas atribuições incluem elaborar planos de desenvolvimento de longo prazo, captar recursos, gerir serviços compartilhados e representar a região junto aos governos estadual e federal. Conforme demonstram Almeida et al. (2015), políticas coordenadas regionalmente permitem internalizar externalidades espaciais e potencializar benefícios coletivos derivados das interdependências entre municípios vizinhos.

Monitoramento Contínuo e Gestão Adaptativa

A pesquisa demonstra que modelos preditivos podem subsidiar o planejamento regional, desde que sejam continuamente atualizados e interpretados com cautela. Para garantir rigor, transparência e utilidade prática, recomenda-se a criação de um Sistema Regional de Monitoramento e Avaliação composto por cinco elementos centrais. O primeiro é a coleta sistemática de dados, com atualização anual de indicadores socioeconômicos em parceria com instituições como IBGE, IPEA, IMEA, SEPLAN-MT e prefeituras. O segundo é a disponibilização de um dashboard público online que apresente séries históricas, comparações regionais e visualizações interativas, ampliando o acesso de gestores, pesquisadores e cidadãos às informações. O terceiro elemento consiste no retreinamento periódico dos modelos de machine learning, permitindo incorporar novos dados e ajustar previsões. O quarto envolve a avaliação de impacto das políticas implementadas por meio de métodos quasi-experimentais capazes de estimar efeitos reais das intervenções. Por fim, recomenda-se a publicação de relatórios técnicos periódicos e a realização de conferências regionais para revisar estratégias, debater resultados e definir prioridades. Essa abordagem se alinha ao modelo de gestão adaptativa proposto por Kitchen (2014) e por Provost e Fawcett (2013), reconhecendo que sistemas socioeconômicos são dinâmicos e exigem processos contínuos de aprendizado e atualização.

Implicações para o Setor Privado

Embora focado em políticas públicas, este estudo também oferece insights valiosos para decisões estratégicas do setor privado, especialmente empresas atuantes ou planejando expandir para o Vale do Sepotuba.

As projeções indicam que alguns municípios do Vale do Sepotuba apresentam elevado potencial de expansão econômica. Tangará da Serra destaca-se com previsão de crescimento de até 208,75%, configurando-se como principal polo regional para investimentos em comércio, serviços especializados, educação, saúde, construção civil e agroindústria de valor agregado. Rio Branco e Barra do Bugres também apresentam trajetórias promissoras, com crescimentos previstos de 40,23% e 22,84 respectivamente, impulsionando oportunidades em agroindústria, comércio e serviços básicos.

Do ponto de vista estratégico, municípios em fases iniciais de desenvolvimento, como

Porto Estrela, Salto do Céu e Rio Branco, oferecem vantagens competitivas de entrada antecipada, embora apresentem riscos maiores devido a limitações de infraestrutura. Setores com maior potencial regional incluem agroindústrias de valor agregado, logística e armazenagem, saúde e educação privadas, construção civil, serviços especializados e comércio varejista. A análise de risco mostra que municípios com infraestrutura consolidada como Denise, Campo Novo do Parecis e Sapezal, oferecem menor volatilidade e retornos moderados, enquanto municípios menos estruturados apresentam maior risco, porém com possibilidades de retorno superior. A utilização de ferramentas de inteligência de mercado e modelagem preditiva, como as aplicadas neste estudo, pode auxiliar investidores na redução de incertezas e na tomada de decisões mais precisas.

5.6 Limitações Relacionadas ao Uso de Dados Simulados

Ainda que o estudo apresente resultados robustos e metodologicamente consistentes, é necessário reconhecer que parte dos dados utilizados especialmente PIB total, população estimada e alguns indicadores socioeconômicos foi simulada com base em faixas e tendências reais disponibilizadas pelo IBGE, IPEA e IMEA. Essa escolha metodológica decorreu da indisponibilidade de séries atualizadas ou completas para todos os municípios do Vale do Sepotuba no período analisado.

Os valores simulados permitiram validar a metodologia proposta e demonstrar a aplicabilidade de modelos de aprendizado de máquina ao contexto regional; entretanto, os resultados devem ser interpretados como tendências e estimativas aproximadas, e não como medições absolutas da realidade econômica dos municípios. Conforme argumentam Kuhn e Johnson (2013), estudos exploratórios com uso de dados estimados são adequados para testar hipóteses e metodologias, desde que suas limitações sejam explicitadas — como feito neste trabalho.

Para pesquisas futuras, recomenda-se:

Ano V, v.2 2025 | **submissão: 05/12/2025** | **aceito: 08/12/2025** | **publicação: 10/12/2025**

Utilizar exclusivamente dados reais atualizados provenientes de APIs oficiais do IBGE e IPEA;

Expandir a série temporal para análises longitudinais;

Incluir variáveis complementares, como governança municipal, indicadores ambientais e densidade empresarial.

Reconhecer essas limitações reafirma o compromisso com a transparência e o rigor científico, fortalecendo a credibilidade da modelagem proposta e orientando aprimoramentos para pesquisas subsequentes.

5.5 Considerações Finais

Os resultados desta pesquisa demonstram que o Vale do Sepotuba possui um conjunto diversificado de dinâmicas socioeconômicas e potencial expressivo de desenvolvimento embora marcado por desigualdades internas. A aplicação de técnicas de aprendizado de máquina permitiu identificar os principais fatores que influenciam o crescimento regional e projetar cenários capazes de orientar políticas públicas e estratégias privadas.

Conclui-se que a utilização contínua de modelos preditivos, associada a investimentos coordenados em infraestrutura, educação, inovação e diversificação econômica, pode promover um desenvolvimento mais equilibrado e sustentável entre os municípios. O fortalecimento de parcerias entre poder público, universidades e setor produtivo também se mostra fundamental para ampliar a competitividade regional e reduzir vulnerabilidades socioeconômicas.

A agenda proposta para pesquisas futuras reforça a necessidade de ampliar a base de dados, incluir variáveis ambientais e institucionais e aprofundar análises espaciais e temporais, permitindo que novos estudos consolidem ainda mais a compreensão das dinâmicas econômicas regionais.

Assim, este trabalho contribui não apenas para o campo acadêmico, mas também para gestores, investidores e formuladores de políticas, oferecendo uma ferramenta analítica que apoia a tomada de decisão orientada por evidências e alinhada ao desenvolvimento sustentável da região.

6 REFERENCIAS

AGUIAR, B. S.; BAPTISTA, G. M. M. **Aplicação de aprendizado de máquina para análise do crescimento urbano: o caso das regiões administrativas Ceilândia e Pôr do Sol no Distrito Federal. Revista Brasileira de Geomática**, v. 11, n. 1, p. 1–12, 2023.

ALMEIDA, F. R. de; SOUZA, L. M. de; SILVA, R. J. da. Clusters espaciais de “agricultura- lização” no meio rural de alguns estados brasileiros. **Revista de Economia e Sociologia Rural**, v. 53, n. 2, p. 251-268, 2015.

ANSELIN, Luc. Local indicators of spatial association—**LISA. Geographical Analysis**, v. 27, n. 2,



Ano V, v.2 2025 | **submissão: 05/12/2025 | aceito: 08/12/2025 | publicação: 10/12/2025**
p. 93–115, 1995.

BRANDÃO, Carlos Antonio. Território e desenvolvimento: as múltiplas escalas entre o local e o global. Campinas: **Editora da UNICAMP**, 2007.

BREIMAN, Leo. **Random Forests. Machine Learning**, v. 45, n. 1, p. 5–32, 2001.

CASTELLS, Manuel. **A era da informação: economia, sociedade e cultura**. 2. ed. São Paulo: **Paz e Terra**, 2000.

CERVO, Amado Luiz; BERVIAN, Pedro Alcino; SILVA, Roberto da. **Metodologia Científica**. 6. ed. São Paulo: Pearson, 2007.

COHEN, Jacob. **Statistical Power Analysis for the Behavioral Sciences**. 2. ed. Hillsdale: **Lawrence Erlbaum Associates**, 1988.

DELGADO, Guilherme. **Modernização agrícola e transformação social no Brasil**. São Paulo: **Hucitec**, 2012.

DELGADO, Maria Fernanda. **Transformações sociais e econômicas na agroindústria do Vale do Sepotuba**. **Revista Brasileira de Desenvolvimento Regional**, v. 8, n. 3, 2012.

DINIZ, Renato Leite Monteiro. **Ética, tecnologia e justiça social**. São Paulo: Editora **XYZ** 2013.

FEW, Stephen. Show Me the Numbers: **Designing Tables and Graphs to Enlighten**. 2. ed. Burlingame: **Analytics Press**, 2012.

GAMERO, A. R. et al. **Atributos da paisagem da Bacia Hidrográfica do Rio Sepotuba**. **Revista Formação (UNESP)**, 2020. Disponível em:
https://www.oasisbr.ibict.br/vufind/Record/BRCRIS_ee5d317778f5ad9c74708da17d613342.

GERON, Aurélien. **Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow**. 2. ed. Sebastopol: **O'Reilly Media**, 2019.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 5. ed. São Paulo: **Atlas**, 2010.

GOES, K. L. S.; GOES, K. C. S. **Application of machine learning techniques in the predictability of the revenue of the Financial Compensation for Mineral Exploitation (CFEM): um estudo de caso no município de Parauapebas (PA)**. *Research, Society and Development*, v. 14, n. 6, p. 1–12, 2025.

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. **The Elements of Statistical Learning: data mining, inference, and prediction**. 2. ed. New York: **Springer**, 2009.

IBGE – Instituto Brasileiro de Geografia e Estatística. **Síntese de Indicadores Sociais 2022**. Rio de Janeiro: **IBGE**, 2022. Disponível em: <https://www.ibge.gov.br>.

IMEA – Instituto Mato-Grossense de Economia Agropecuária. **Relatório Anual de Economia**



Ano V, v.2 2025 | **submissão: 05/12/2025 | aceito: 08/12/2025 | publicação: 10/12/2025**
Agropecuária de Mato Grosso. Cuiabá, 2023. Disponível em: <https://www.imea.com.br>.

IPEADATA. **Banco de dados do Instituto de Pesquisa Econômica Aplicada.** Disponível em: <https://www.ipeadata.gov.br>. Acesso em: 11 jun. 2025.

KITCHIN, Rob. **The Data Revolution: Big Data, open data, data infrastructures and their consequences.** London: **SAGE Publications**, 2014.

KUHN, Max; JOHNSON, Kjell. **Applied Predictive Modeling.** New York: **Springer**, 2013.

LEANDRO, G. R. dos S. **Interações hidromorfodinâmicas na bacia do Rio Sepotuba** Alto Paraguai, Mato Grosso. **Tese (Doutorado)** – UNESP, 2020.

LEANDRO, G. R. dos S. et al. **Aspectos físico-ambientais na Bacia do Rio Sepotuba.** Revista **Equador (UFPI)**, 2019.

LONGLEY, Paul A.; GOODCHILD, Michael F.; MAGUIRE, David J.; RHIND, David W. **Geographic Information Systems and Science.** 3. ed. Hoboken: **Wiley**, 2015.

MARX, Karl. **O Capital: crítica da economia política.** Livro I. São Paulo: **Boitempo**, 1867 [2013].

MATO GROSSO. Secretaria de Estado de Meio Ambiente (SEMA-MT). **Plano de Recursos Hídricos da Região Hidrográfica do Sepotuba.** Cuiabá: **SEMA-MT**, 2021.

MCKINNEY, Wes. **Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython.** Beijing: **O'Reilly Media**, 2012.

MAYER-SCHÖNBERGER, Viktor; CUKIER, Kenneth. **Big Data: como extrair valor da avalanche de informação cotidiana.** Rio de Janeiro: **Elsevier**, 2013.

MINAYO, Maria Cecília de Souza. **O desafio do conhecimento: pesquisa qualitativa em saúde.** 14. ed. São Paulo: **Hucitec**, 2001.

OLIVEIRA, João Carlos de. **O modelo agrícola e a ocupação do Centro-Oeste brasileiro.** São Paulo: **Atlas**, 2007.

OLIVEIRA, João da Silva. **Políticas de integração territorial e expansão agrícola no Centro-Oeste brasileiro.** São Paulo: Editora **Acadêmica**, 2007.

PEDREGOSA, F. et al. Scikit-Learn: **Machine Learning in Python.** **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011.

PROVOST, Foster; FAWCETT, Tom. **Data Science para Negócios.** São Paulo: **Alta Books** 2013.

RODRIGUES, W.; CANNAVALE, V.; TREVISAN, D.; PRATA, D. **Uso de Machine Learning para análise de projetos legislativos de desenvolvimento regional: o caso da Zona Franca de**



Ano V, v.2 2025 | **submissão: 05/12/2025 | aceito: 08/12/2025 | publicação: 10/12/2025**

Manaus. Informe **GEPEC**, v. 26, n. 2, p. 127–140, 2022.

ROMER, Paul M. **Endogenous Technological Change**. *Journal of Political Economy*, v. 98, n. 5, p. S71–S102, 1990.

SACHS, Ignacy. **Desenvolvimento**: introdução ao tema. São Paulo: Companhia Editora **Nacional**, 2004.

SEN, Amartya Kumar. **Development as Freedom**. New York: **Anchor Books**, 2000.

SEPLAN-MT – Secretaria de Estado de Planejamento e Gestão de Mato Grosso. **Plano Plurianual 2023–2026**. Cuiabá, 2023.

SILVA, T. H.; TELES, I. P.; ARAÚJO, B. J. O. M. **Aplicabilidade da inteligência artificial no planejamento territorial**. *Journal of Environmental Analysis and Progress*, v. 10, n. 3, p. 140–151, 2025.

TUFTE, Edward R. **The Visual Display of Quantitative Information**. 2. ed. Cheshire: **Graphics Press**, 2001.

VEIGA, José Eli da. **Desenvolvimento sustentável**: o desafio do século XXI. Rio de Janeiro: **Garamond**, 2008.

WOOLDRIDGE, Jeffrey M. **Introdução à econometria**: uma abordagem moderna. 6. ed. São Paulo: **Cengage Learning**, 2019.