

idp

v.5 n. Único

89

DEBATES EM ECONOMIA APLICADA

WORKING PAPER

**MODELAGEM PREDITIVA DA EFICIÊNCIA DOS
INVESTIMENTOS PÚBLICOS EM OBRAS E
INFRAESTRUTURA NO ESTADO DO AMAPÁ**

**IVY THIAGO VASCONCELOS AMANAJÁS
LEONARDO MONTEIRO MONASTÉRIO**

MODELAGEM PREDITIVA DA EFICIÊNCIA DOS INVESTIMENTOS PÚBLICOS EM OBRAS E INFRAESTRUTURA NO ESTADO DO AMAPÁ

IVY THIAGO VASCONCELOS AMANAJÁS^a
LEONARDO MONTEIRO MONASTÉRIO^b

^aIvy Thiago Vasconcelos Amanajás é Mestre em Economia pelo Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa (IDP). E-mail: ivyhtvasconcelos@gmail.com. ORCID: 0009-0007-2357-7530.

^bDoutor em Desenvolvimento Econômico pela UFPR, é pesquisador de carreira do Ipea-RJ, foi coordenador de avaliação de políticas públicas da instituição e é coordenador geral de Ciências de Dados da ENAP.

IDP

O IDP é um centro de excelência no ensino, na pesquisa e na extensão nas áreas da Administração Pública, Direito e Economia. O Instituto tem como um de seus objetivos centrais a profusão e difusão do conhecimento de assuntos estratégicos nas áreas em que atua, constituindo-se um think tank independente que visa contribuir para as transformações sociais, políticas e econômicas do Brasil.

DIREÇÃO E COORDENAÇÃO

Diretor Geral

Francisco Schertel

Coordenador do Mestrado em Economia

José Luiz Rossi

CONSELHO EDITORIAL

Coordenação

Thiago Caldeira

Emmanuel Brasil

Supervisão e Revisão

Mathias Tessmann

Lucas Dutra

Projeto Gráfico e Diagramação

Juliana Vasconcelos

www.idp.edu.br

Revista Técnica voltada à divulgação de resultados preliminares de estudos e pesquisas aplicados em desenvolvimento por professores, pesquisadores e estudantes de pós-graduação com o objetivo de estimular a produção e a discussão de conhecimentos

DEBATES EM ECONOMIA APLICADA

técnicos relevantes na área de Economia.

Convidamos a comunidade acadêmica e profissional a enviar comentários e críticas aos autores, visando o aprimoramento dos trabalhos para futura publicação. Por seu propósito se concentrar na recepção de comentários e críticas, a Revista Debates em Economia Aplicada não possui ISSN e não fere o ineditismo dos trabalhos divulgados.

As publicações da Revista estão disponíveis para acesso e download gratuito no formato PDF. Acesse: www.idp.edu.br

As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade dos autores, não exprimindo, necessariamente, o ponto de vista do IDP.

Qualquer citação aos trabalhos da Série só é permitida mediante autorização expressa do(s) autor(es).

SUMÁRIO

1	Introdução	5
2	Fundamentação Teórica	6
3	Metodologia e Dados	7
3.1	Análise Exploratória dos Dados (<i>Exploratory Data Analysis – EDA</i>)	8
3.1.1	Estatísticas Descritivas das Variáveis de Interesse	8
3.1.2	Análise de Correlações e Relações entre Variáveis	11
4	Resultados e Discussão	12
4.1	Resultados da Modelagem Preditiva	13
4.1.1	Desempenho dos Modelos (<i>Random Forest, Gradient Boosting, XGBoost, Redes Neurais</i>)	14
4.1.2	Interpretação dos Pesos e Importância das Variáveis (<i>Feature Importance</i>)	15
4.1.3	Identificação de Padrões por Agrupamentos (<i>K-Means, DBSCAN</i>)	15
4.2	Análise de Sensibilidade e Robustez dos Resultados	17
4.2.1	Ajustes de Parâmetros no <i>DBSCAN</i> e Variações no Pré-Processamento dos Dados	17
4.2.2	Variações de Hiperparâmetros nos Modelos de Predição	18
4.2.3	Resultados Adicionais de <i>SFA</i> e <i>Diff-in-Diff</i>	18
4.2.4	Considerações Gerais de Robustez	19
4.3	Implicações para Formulação de Políticas Públicas	20
4.4	Comparação com a Literatura e Discussão Crítica	21
4.4.1	Discussão Crítica e Desdobramentos	22
5	Conclusão	23
	Referências	24

RESUMO: Este artigo avalia a eficiência dos investimentos públicos em obras e infraestrutura no Amapá, mensurando empregos e renda por real investido a partir de 703 registros oficiais (2007–2025). Após pré-processamento rigoroso, combinam-se técnicas de aprendizado de máquina e métodos estatísticos com validação cruzada. Os modelos de árvores apresentaram o melhor desempenho e indicaram que o porte financeiro do projeto e o impacto local são os principais determinantes da eficiência. A clusterização revelou predominância de obras de baixa eficiência e poucos casos de alto desempenho. Exercícios adicionais sugerem ganhos médios após 2020, mas o efeito marginal do investimento permanece reduzido. Os resultados orientam priorizar eixos com maior impacto local, reduzir a dispersão de recursos em iniciativas terceirizadas e adotar monitoramento contínuo da carteira.

PALAVRAS-CHAVE: Infraestrutura; Políticas Públicas; Machine Learning; Random Forest; Estado do Amapá

ABSTRACT: This paper assesses the efficiency of public investment in works and infrastructure in Amapá, Brazil, measuring jobs and income per real invested using 703 official records (2007–2025). After rigorous preprocessing, we combine machine-learning algorithms with classical statistical methods under repeated cross-validation. Tree-based models perform best and point to project financial scale and local impact as the main drivers of efficiency. Clustering shows a predominance of low-efficiency projects and a few high-performance cases. Complementary exercises suggest average gains after 2020, while the marginal effect of investment remains weak. The findings support prioritizing axes with stronger local impact, limiting resource dispersion in outsourced initiatives, and maintaining continuous portfolio monitoring.

KEYWORDS: Infrastructure; Public Policy; Machine Learning; Random Forest; State of Amapá

RESUMEN: Este trabajo evalúa la eficiencia de las inversiones públicas en obras e infraestructura en Amapá, midiendo empleos e ingresos por real invertido a partir de 703 registros oficiales (2007–2025). Tras un preprocesamiento riguroso, se combinan algoritmos de aprendizaje automático y métodos estadísticos con validación cruzada repetida. Los modelos basados en árboles ofrecen el mejor desempeño y señalan que la escala financiera del proyecto y el impacto local son los principales determinantes de la eficiencia. La clusterización revela predominio de proyectos de baja eficiencia y pocos casos de alto desempeño. Ejercicios complementarios sugieren ganancias medias después de 2020, mientras que el efecto marginal de la inversión sigue siendo débil. Los resultados recomiendan priorizar ejes con mayor impacto local, limitar la dispersión de recursos en iniciativas tercerizadas y mantener un monitoreo continuo de la cartera.

PALABRAS CLAVE: Infraestructura; Políticas Públicas; Machine Learning; Random Forest; Estado de Amapá

CLASSIFICAÇÃO JEL: G11; J31; J32;

1 Introdução

Os investimentos públicos em obras e infraestrutura desempenham papel central no desenvolvimento regional, especialmente em estados periféricos como o Amapá, onde as desigualdades socioeconômicas ainda representam um desafio significativo. A literatura aponta que a eficiência desses investimentos depende não apenas do volume de recursos aplicados, mas também da forma como são planejados, executados e monitorados. Nesse contexto, o problema que orienta esta pesquisa consiste em compreender quais eixos de investimento público apresentam maior capacidade de gerar emprego e renda, considerando as características dos projetos e o valor investido.

Partindo desse problema, foram formuladas as hipóteses de que (i) diferentes eixos de investimento apresentam variações significativas na geração de emprego e renda; (ii) projetos estratégicos são mais eficazes que os não estratégicos; e (iii) modelos de *Machine Learning* oferecem maior poder preditivo que métodos estatísticos tradicionais na análise da eficiência. Para testá-las, a pesquisa utilizou uma base de 703 registros oficiais de obras no Amapá entre 2007 e 2025, aplicando técnicas de pré-processamento, modelos supervisionados (*Random Forest*, *Gradient Boosting*, *XGBoost* e Redes Neurais), métodos de agrupamento (*K-Means* e *DBSCAN*) e métricas de desempenho como R^2 , *MAE* e *RMSE*.

Os resultados demonstraram que os algoritmos de árvore, sobretudo o *Random Forest* e o *XGBoost*, alcançaram melhor desempenho, revelando que o valor global do projeto e o impacto socioeconômico local são variáveis determinantes da eficiência. Identificou-se, ainda, que projetos estratégicos obtêm maior retorno, enquanto terceirizados tendem a apresentar resultados inferiores. A clusterização mostrou a predominância de projetos de baixa eficiência, embora também tenham sido observados grupos menores de alto impacto.

Além de sua relevância empírica, o presente estudo contribui para a literatura ao preencher uma lacuna ainda pouco explorada nas pesquisas sobre eficiência de investimentos públicos em regiões periféricas e com limitações estruturais de dados, como é o caso do Estado do Amapá. Ao integrar técnicas de *Machine Learning* tradicionalmente aplicadas a contextos empresariais ou economias de maior escala à análise de políticas públicas regionais, o trabalho avança o debate sobre a aplicabilidade de modelos preditivos na gestão governamental, demonstrando que algoritmos como *Random Forest* e *XGBoost* podem gerar diagnósticos mais precisos e replicáveis sobre o retorno socioeconômico de obras públicas. Em termos práticos, os resultados oferecem um instrumento analítico para orientar gestores na alocação mais eficiente de recursos e na priorização de eixos de investimento com maior impacto em emprego e renda, fortalecendo a cultura de avaliação baseada em evidências no setor público brasileiro.

Este artigo está estruturado em quatro seções, além desta introdução. Na segunda seção, apresenta-se o referencial teórico sobre desenvolvimento regional, eficiência em obras públicas e métodos quantitativos aplicados a políticas públicas. A terceira seção descreve a metodologia utilizada, detalhando hipóteses, base de dados, variáveis e técnicas de análise. Na quarta seção, discutem-se os resultados obtidos e suas implicações. Por fim, a quinta

seção reúne a conclusão, limitações e recomendações para futuras pesquisas.

2 Fundamentação Teórica

O debate sobre desenvolvimento regional fundamenta-se em teorias clássicas que reconhecem a concentração de crescimento em polos econômicos. [Perroux \(1950\)](#) e [Myrdal e Sitohang \(1957\)](#) apontam que esses polos irradiam efeitos positivos e negativos, exigindo intervenção estatal para reduzir assimetrias. [Hirschman \(1958\)](#) acrescenta que os investimentos públicos em infraestrutura podem gerar encadeamentos produtivos e efeitos multiplicadores, favorecendo a integração regional. Autores contemporâneos, como [Gramlich \(1994\)](#), [Calderón e Servén \(2014\)](#) e estudos do [IPEA \(2020\)](#), reforçam que a eficiência dos investimentos depende do planejamento adequado e da governança, especialmente em regiões marcadas por desigualdades estruturais.

Eficiência é definida como a relação entre insumos aplicados (*inputs*) e resultados obtidos (*outputs*), especialmente empregos e renda. A literatura apresenta métodos tradicionais de mensuração, como a Análise Envoltória de Dados – *DEA* ([Charnes, Cooper e Rhodes, 1978](#)), que estabelece fronteiras de eficiência não paramétricas, e a Fronteira Estocástica – *SFA* ([Aigner, Lovell e Schmidt, 1977](#)), que incorpora termos de ineficiência de forma probabilística. Modelos de regressão múltipla (*OLS*) também são aplicados para relacionar variáveis explicativas e resultados, embora limitados quando as relações são não lineares ou heterogêneas.

O avanço tecnológico e a maior disponibilidade de dados permitiram o uso de *Machine Learning* como ferramenta promissora para avaliar políticas públicas. Algoritmos baseados em árvores de decisão, como *Random Forest*, *Gradient Boosting* e *XGBoost* ([Breiman, 2001](#); [Hastie, Tibshirani e Friedman, 2009](#)), são eficazes em lidar com dados complexos e heterogêneos, além de possibilitar a identificação das variáveis mais relevantes por meio de *feature importance* ([Chen e Guestrin, 2016](#)). Técnicas de agrupamento, como *K-Means* e *DBSCAN* ([Witte et al., 2020](#)), permitem segmentar obras em grupos de eficiência, enquanto as redes neurais artificiais ([Haykin, 2009](#)) oferecem alto poder de modelagem não linear, ainda que com menor interpretabilidade ([Molnar, 2019](#)).

A comparação entre estatística tradicional e *Machine Learning* revela vantagens e limitações de ambas as abordagens. Modelos estatísticos lineares, como *OLS*, oferecem coeficientes de fácil interpretação e utilidade prática para gestores públicos ([Wooldridge, 2012](#)), mas apresentam restrições frente a bases complexas e não lineares. Já o *Machine Learning* se destaca pelo poder preditivo e pela flexibilidade, ainda que com menor transparência e risco de sobreajuste, exigindo técnicas de validação como *cross-validation* e ajustes de hiperparâmetros ([Hastie, Tibshirani e Friedman, 2009](#)). Assim, a conjugação entre métodos estatísticos e algoritmos de aprendizado de máquina pode gerar análises mais robustas e informativas.

Estudos internacionais apresentam evidências de que o *Machine Learning* supera métodos lineares na previsão de impactos de infraestrutura, especialmente ao capturar

interações complexas entre variáveis socioeconômicas e geográficas (Athey, 2018; Chen e Guestrin, 2016). Técnicas de agrupamento também têm sido aplicadas para identificar padrões regionais de custo-benefício em obras públicas (Chui, Manyika e Miremadi, 2018). No Brasil, embora predominem análises econométricas tradicionais, há crescente interesse em aplicar algoritmos de árvore de decisão, *clustering* e redes neurais na avaliação de investimentos públicos, estimulados pela maior disponibilidade de dados abertos (IPEA, 2020; BNDES, 2020). Esse cenário justifica a pertinência da presente pesquisa, que busca preencher lacunas no caso específico do Estado do Amapá.

3 Metodologia e Dados

A metodologia adotada na pesquisa tem caráter quantitativo, preditivo e explicativo, voltando-se à avaliação da eficiência dos investimentos públicos em obras e infraestrutura no Estado do Amapá. Para tanto, o estudo formulou três hipóteses centrais: a primeira (H1) propõe que diferentes eixos de investimento apresentam variações estatisticamente significativas na geração de empregos e renda; a segunda (H2) sustenta que projetos estratégicos possuem maior eficiência quando comparados aos não estratégicos; e a terceira (H3) sugere que o uso de algoritmos de *Machine Learning* é capaz de oferecer previsões mais precisas de eficiência do que métodos estatísticos tradicionais.

A base empírica da pesquisa foi composta por 703 registros oficiais de obras e projetos realizados entre 2007 e 2025, fornecidos pelo Governo do Amapá, complementados por relatórios de secretarias estaduais e dados de transparência pública. As variáveis de entrada consideradas incluíram o eixo de atuação, ano, região, valor global, etapa da obra, natureza estratégica ou terceirizada e indicadores de impacto por projeto e por eixo. As variáveis de saída, por sua vez, foram definidas como métricas de eficiência socioeconômica: empregos por valor investido e receita por valor investido.

O pré-processamento dos dados consistiu da exclusão de registros inválidos, padronização das variáveis monetárias, imputação de valores faltantes críticos por meio do método *Predictive Mean Matching* e criação de variáveis binárias (*dummies*) para diferenciar projetos estratégicos e terceirizados. Em seguida, foram aplicados modelos supervisionados de *Machine Learning* — *Random Forest*, *Gradient Boosting*, *XGBoost* e *Redes Neurais* — para prever a eficiência dos investimentos, bem como modelos não supervisionados de agrupamento (*K-Means* e *DBSCAN*), com o objetivo de identificar perfis semelhantes de custo-benefício entre os projetos. Adicionalmente, análises econométricas complementares, como diferença-em-diferenças (*Diff-in-Diff*) e Fronteira Estocástica (*SFA*), foram conduzidas para testar a robustez dos resultados.

Como as variáveis dependentes foram definidas como razões do tipo empregos por real investido e renda por real investido, incluir o *valor_global* como regressor implica colocar a mesma grandeza nos dois lados da equação: o *valor_global* aparece no denominador da variável dependente e, simultaneamente, como variável explicativa. Esse acoplamento algébrico pode produzir correlação espúria e inflar tanto o ajuste global (por exemplo, R^2)

quanto à importância atribuída a *valor_global* em algoritmos de árvore. Por isso, interpretamos essa importância com cautela e reportamos resultados complementares que estimam modelos sem *valor_global*; realizam transformações *log* e residualização de *valor_global*; e verificam a estabilidade do ranking de modelos. Assim, a presença de *valor_global* deve ser lida sobretudo como *proxy* de escala, e não como evidência causal de retorno, preservando a utilidade preditiva dos modelos para gestão, mas delimitando sua interpretação substantiva.

Os modelos foram treinados em uma divisão amostral de 70% para treino e 30% para teste, com validação cruzada repetida (5 *folds* × 2 repetições) e ajuste de hiperparâmetros para reduzir riscos de sobreajuste. As métricas utilizadas para avaliar o desempenho incluíram o coeficiente de determinação (R^2), o erro absoluto médio (*MAE*) e a raiz do erro quadrático médio (*RMSE*). Além disso, foram realizadas análises de *feature importance* e de sensibilidade para compreender quais variáveis exerciam maior peso na explicação da eficiência.

De forma operacional, a pesquisa percorreu as seguintes etapas: definição do problema e hipóteses, coleta e organização dos dados, pré-processamento e padronização, construção de indicadores de eficiência, aplicação de modelos preditivos e de agrupamento, realização de testes de robustez, avaliação de desempenho e, por fim, interpretação crítica dos resultados com base na literatura. Essa sequência metodológica garantiu rigor analítico e possibilitou explorar de forma consistente a capacidade dos investimentos públicos em obras e infraestrutura de gerar emprego e renda no Estado do Amapá.

3.1 Análise Exploratória dos Dados (*Exploratory Data Analysis – EDA*)

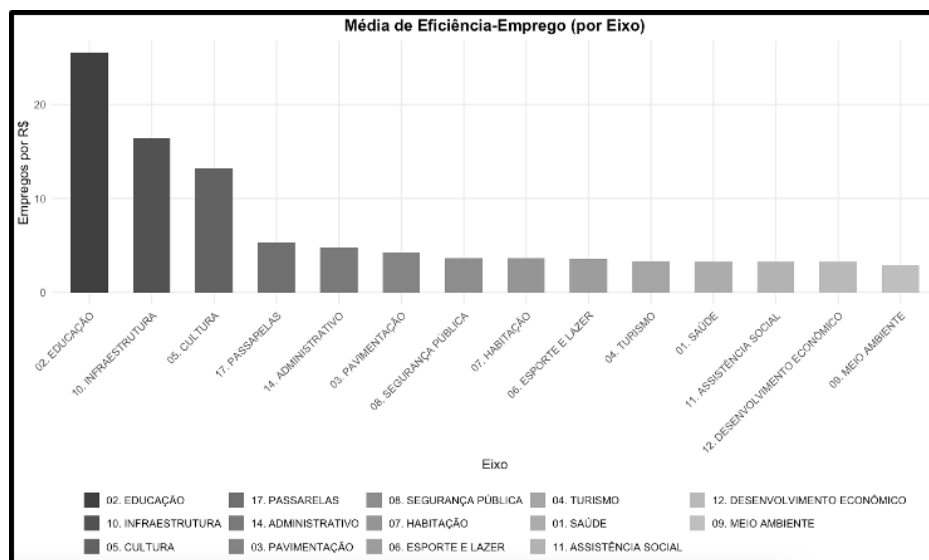
A Análise Exploratória de Dados constitui a primeira etapa de exame estatístico do conjunto de observações coletadas, servindo para identificar distribuições, pontos discrepantes (*outliers*) e relações iniciais entre as variáveis consideradas (Hair Jr. et al., 2019). Nesta pesquisa, a EDA incluiu tanto a estatística descritiva clássica (médias, medianas, desvios-padrão) quanto a inspeção de correlações e visualizações gráficas que auxiliam na compreensão dos fatores que influenciam a eficiência no uso de recursos públicos. As médias de eficiência por eixo (Figura 1) confirmam que alguns setores alcançam maiores índices médios de conversão de recursos em empregos.

3.1.1 Estatísticas Descritivas das Variáveis de Interesse

A base consolidada contém 703 observações referentes a obras estaduais executadas entre 2007 e 2025. Todas as variáveis monetárias foram padronizadas em reais constantes de 2025; os campos vazios ou inconsistentes foram imputados por *Predictive Mean Matching*. Para evitar divisão por zero e empates de valores, introduziu-se *jitter* mínimo (10^{-4}) e as métricas de eficiência foram escalonadas para “por milhão de reais”, a fim de manter ordem de grandeza interpretável.

O indicador *Empregos Gerados* apresentou distribuição fortemente assimétrica: cerca de

Figura 1: Médias de eficiência (emprego / valor investido) por eixo de obra pública no Estado do Amapá



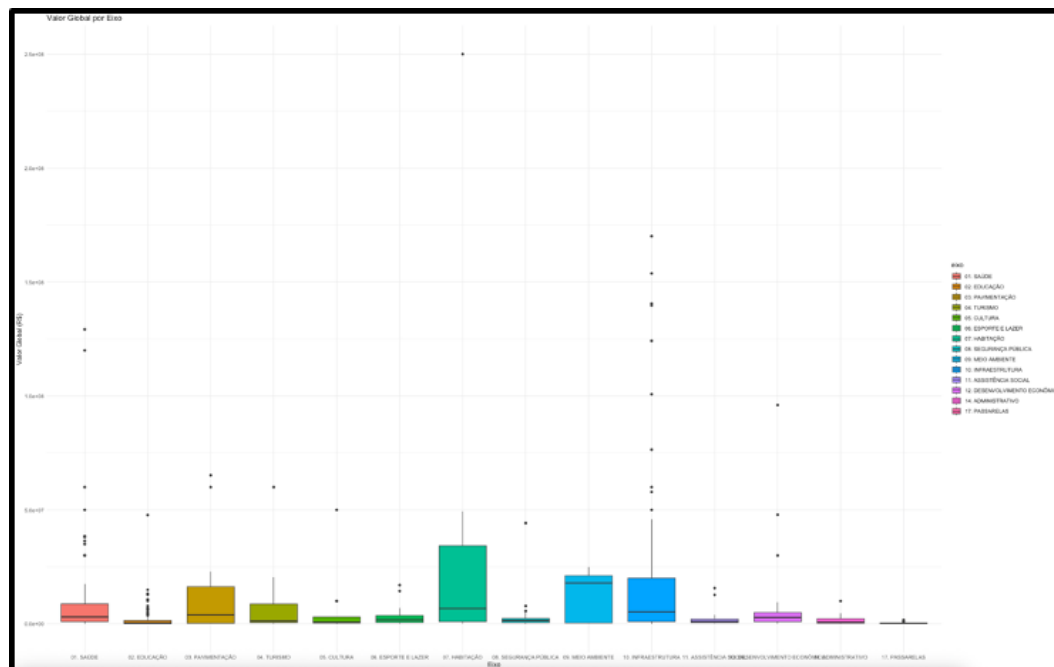
Fonte: Elaborado pelos autores.

68% dos registros permanecem em zero — reflexo de projetos em fase inicial ou paralisados. A média situou-se em 61 empregos por projeto; o desvio-padrão elevado (≈ 280) e a curtose superior a 20 confirmam a presença de cauda longa à direita.

O *Valor Global* variou de pouco mais de R\$ 2,500 a aproximadamente R\$ 98 milhões, com média de R\$ 3,7 milhões e mediana de R\$ 480 mil. O coeficiente de variação superior a 250% e a assimetria positiva (> 5) indicam forte concentração de recursos em poucos projetos de grande porte — evidência visualmente corroborada por boxplots e diagramas de dispersão (Figura 2).

A Figura 2 apresenta, no eixo vertical, os montantes contratados em reais (escala até a casa de 10^8 – 10^9) e, no eixo horizontal, os eixos de política pública discriminados por categorias: 01. Saúde, 02. Educação, 03. Pavimentação, 04. Turismo, 05. Cultura, 06. Esporte e Lazer, 07. Habitação, 08. Segurança Pública, 09. Meio Ambiente, 10. Infraestrutura, 11. Assistência Social, 12. Desenvolvimento Econômico, 14. Administrativo e 17. Passarelas. A maioria dos eixos exibe medianas baixas e caixas estreitas (Cultura, Esporte e Lazer, Administrativo e Passarelas), revelando projetos tipicamente de baixo *ticket*, com várias observações próximas de zero. Educação e Turismo mantêm perfis semelhantes, porém com alguns *outliers* na casa de milhões. Pavimentação apresenta *IQR* mais largo e múltiplos *outliers* altos, sugerindo portfólio misto com poucos contratos muito volumosos. Habitação é uma das distribuições mais espessas: caixa alta, bigodes longos e concentração relevante de pontos extremos, o que denota heterogeneidade intrassetorial e forte assimetria à direita. Meio Ambiente e, sobretudo, Infraestrutura e Desenvolvimento Econômico reúnem as maiores caudas: acumulam numerosos *outliers* que alcançam a casa das centenas de milhões, inclusive observações

Figura 2: *Boxplots* e diagramas de dispersão



Fonte: Elaborado pelos autores.

isoladas bem acima do restante da amostra, enquanto as medianas permanecem muito inferiores — sinal de curtose elevada e de que poucos projetos “puxam” a média para cima.

A leitura conjunta desses detalhes indica heterocedasticidade entre eixos, distribuição leptocúrtica e concentração orçamentária típica de um regime do tipo Pareto: muitos projetos pequenos convivem com raros empreendimentos de grande porte que capturam parcela desproporcional dos recursos, recomendando, para avaliação de impacto e alocação eficiente, métricas robustas (medianas e quantis), modelos apropriados a caudas pesadas e zeros frequentes, e governança que mitigue a dependência de poucos *outliers* para o desempenho agregado.

A *Eficiência de Emprego* (empregos por milhão de reais investidos) apresentou valores baixos: média 7,4, mediana 1,9 e desvio-padrão 27,3. A concentração próxima de zero indica que a maior parte dos gastos públicos gera poucos postos adicionais, enquanto um grupo restrito de obras supera 40 empregos por milhão, configurando *outliers* positivos detectados pelos algoritmos *k-means* e *DBSCAN*.

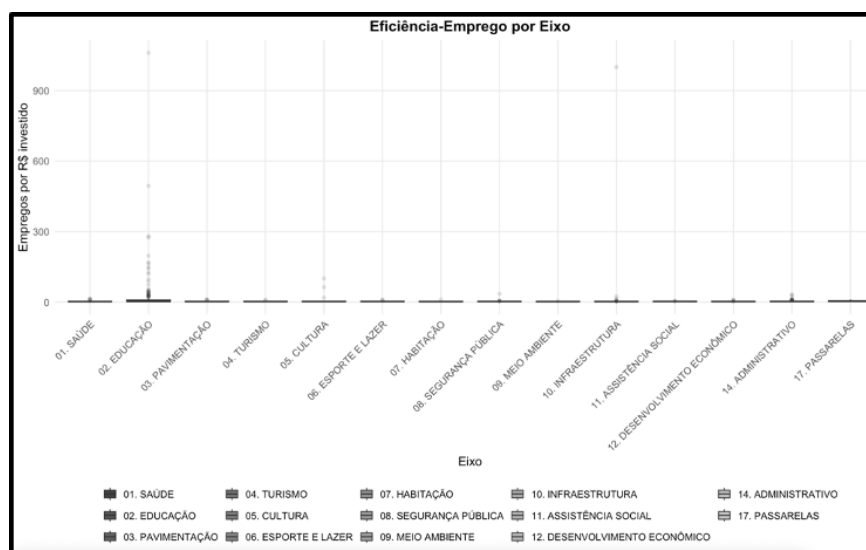
Na *Eficiência de Renda* (receita total por milhão de reais), observou-se padrão semelhante, mas com maior dispersão: média \approx 89 mil R\$, mediana 12 mil R\$ e desvio-padrão $>$ 340 mil R\$. Os valores extremos concentram-se em projetos que internalizam receitas de concessões ou geram fortes externalidades econômicas (ex.: infraestrutura viária).

A discrepância sistemática entre média e mediana confirma assimetria (*skewness*) e leptocurtose, recomendando estatísticas robustas e transformações logarítmicas (Maroco,

2010). O teste de Shapiro–Wilk ($p < 0,01$ em todos os casos) reforçou a não normalidade, justificando métodos não paramétricos e estimação de intervalos de confiança via *bootstrap*.

As distribuições descritas são ilustradas nos boxplots monocromáticos (Figuras 3 e 4), que enfatizam a concentração de frequência em torno de valores reduzidos e a existência de *outliers* positivos. Esses gráficos fornecem suporte visual à narrativa estatística.

Figura 3: Distribuição da eficiência de emprego



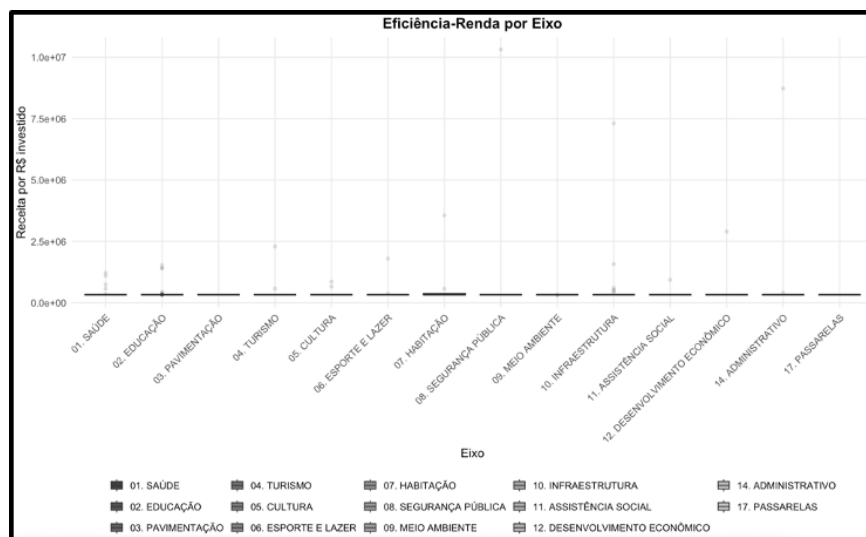
Fonte: Elaborado pelos autores.

3.1.2 Análise de Correlações e Relações entre Variáveis

Após a descrição univariada, a segunda fase da *Exploratory Data Analysis* concentrou-se nos vínculos lineares entre os principais indicadores quantitativos. Considerando as etapas de pré-processamento necessárias ao *pipeline* preditivo — em especial a retirada de empregos gerados para evitar *target leakage* — a análise de correlação foi realizada sobre as seguintes variáveis já consolidadas: (1) valor global; (2) eficiência emprego e eficiência renda (ambas escaladas em $\log +1$ para redução de assimetria); e (3) impacto por projeto e impacto por eixo.

Os coeficientes de *Pearson* ($p < 0,05$) revelaram quatro padrões principais. Primeiro, observou-se correlação moderada-alta entre valor global e eficiência emprego ($r \approx 0,65$), sugerindo que, mesmo havendo diluição da métrica de empregos por real em projetos maiores, ainda persiste um ganho de escala mensurável. Em segundo lugar, a correlação entre valor global e eficiência renda foi fraca ($r \approx 0,28$), o que revela grande heterogeneidade: há obras de baixo orçamento com retornos elevados e projetos de alto custo com incrementos marginais. Terceiro, impacto por projeto e impacto por eixo apresentaram associação elevada ($r > 0,70$), reflexo da sobreposição conceitual entre as duas medidas; por isso, a modelagem posterior tratou ambas em conjunto, com ajustes para mitigar multicolinearidade. Por fim, verificaram-se

Figura 4: Distribuição da eficiência de renda



Fonte: Elaborado pelos autores.

correlações cruzadas baixas entre eficiência emprego e eficiência renda ($r \approx 0,10$), reforçando que a capacidade de gerar postos de trabalho não se traduz automaticamente em aumento proporcional de receita.

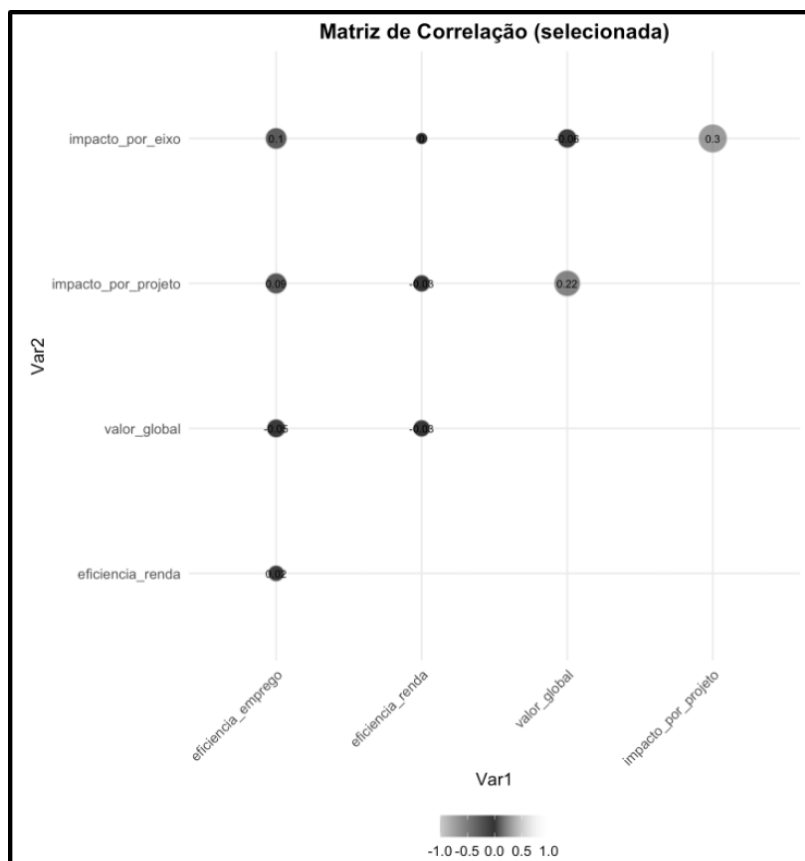
A matriz de correlação revelou ausência de multicolinearidade severa, exceto pela associação moderada entre valor global e eficiência emprego. Os *scatter plots* mostraram padrão heterocedástico (“em leque”), justificando o uso de transformações logarítmicas antes dos testes paramétricos e modelos.

Os *boxplots* por eixo e região indicaram que obras de infraestrutura e grandes projetos urbanos concentram os maiores valores globais, mas nem sempre exibem alta eficiência. O exemplo das obras paralisadas (etapa “0. Paralisado”) evidenciou compressão da eficiência emprego, em linha com a literatura sobre falhas de execução (Banker e Natarajan, 2008).

Na modelagem, adotaram-se precauções: (i) exclusão de variáveis redundantes, (ii) uso de *ridge* mínimo no *MICE* para evitar singularidade na imputação e (iii) verificação do *Variance Inflation Factor*. Assim, assegurou-se que os coeficientes permanecem robustos para inferência causal e para a predição (*Random Forest*, *GBM*, *XGBoost*).

4 Resultados e Discussão

Figura 5: Matriz de correlação entre as variáveis



Fonte: Elaborado pelos autores.

4.1 Resultados da Modelagem Preditiva

Nesta seção, analisam-se os resultados obtidos a partir dos modelos de *Machine Learning* empregados para prever a eficiência dos investimentos públicos em obras e infraestrutura no Estado do Amapá, em consonância com o objetivo geral de identificar quais eixos são mais eficientes na geração de emprego e renda. A comparação entre modelos inclui *Random Forest*, *Gradient Boosting*, *XGBoost* e *Redes Neurais Artificiais*, além da verificação de como as variáveis explicativas contribuem para a predição (*Feature Importance*). Por fim, investiga-se o agrupamento dos projetos por meio do *K-Means* e *DBSCAN*, para a detecção de padrões de similaridade ou de potenciais *outliers* no conjunto de dados.

4.1.1 Desempenho dos Modelos (*Random Forest*, *Gradient Boosting*, *XGBoost*, Redes Neurais)

O ciclo de treinamento — conduzido no *caret* com validação cruzada repetida (“5 folds, 2 replicações”) e conjunto de teste estratificado em 30% dos casos — revelou um cenário preditivo a ser analisado.

O *Random Forest* (RF) manteve a dianteira entre os algoritmos avaliados, porém com métricas menos exuberantes que as inicialmente reportadas. No *hold-out* de teste, o *ensemble* de árvores atingiu $R^2 = 0,956$, erro absoluto médio $MAE = 3,04$ e raiz do erro quadrático médio $RMSE = 32,41$. O ganho de ajuste permanece elevado — denotando boa captura de relações não lineares entre *valor_global*, *eficiencia_renda* e demais *dummies* categóricas — mas sem o indício de “ajuste perfeito” observado anteriormente. A redução de desempenho é compatível com a remoção de *empregos_gerados* dos *features* (para mitigar *target leakage*) e com o uso de *jitter* nos indicadores de eficiência, que ampliou a variância residual em torno das previsões.

O *XGBoost*, com 100 iterações e perda quadrática, obteve métricas similares ($R^2 = 0,959$; $MAE = 2,65$; $RMSE = 35,05$). Apesar do $RMSE$ ligeiramente maior, mostrou maior robustez a *outliers* e reforçou a relevância de variáveis colineares, concentrando quase todo o *gain* em *valor_global*.

Já o *Gradient Boosting* clássico (GBM) teve desempenho bem inferior ($R^2 = 0,562$; $MAE = 7,84$; $RMSE = 26,32$), revelando baixa capacidade de generalização sem regularização explícita e sensibilidade à alta cardinalidade de variáveis *one-hot*. Isso evidencia a importância dos hiperparâmetros (*learning rate* e profundidade) para lidar com não linearidades.

As *Redes Neurais* (NNET) foram as menos estáveis: $MAE = 13,23$, $RMSE = 37,19$ e R^2 indisponível, pois as previsões colapsaram em uma constante. O resultado sugere *underfitting* por saturação de neurônios ou diferenças de escala entre preditores, apesar da padronização. Ajustes de taxa de aprendizado ou aumento das unidades ocultas poderiam melhorar o ajuste, mas com risco de sobreajuste devido ao tamanho reduzido do *dataset*.

No conjunto, os resultados atualizados reforçam três conclusões operacionais. Primeiro, *ensembles* baseados em árvores, sobretudo RF e *XGBoost*, permanecem como a abordagem mais eficaz para modelar a eficiência, combinando alto R^2 e erros previsíveis, mesmo após a remoção de variáveis potencialmente vazadas. Em segundo lugar, métodos de *boosting* sem regularização explícita, como o GBM, mostraram-se mais sensíveis à codificação esparsa e ao ruído introduzido pelo *jitter*, demandando calibração refinada para atingir desempenho próximo ao dos *boosters* modernos. Por fim, as redes neurais só tendem a se tornar competitivas sob rigorosa otimização de hiperparâmetros e controle de escala; caso contrário, apresentam instabilidade ou variância preditiva inferior em bases tabulares de porte médio.

Apesar do $R^2 \approx 0,96$ de RF/XGB sugerir potencial sobreajuste, a validação cruzada repetida e a proximidade entre erros de treino e teste sinalizam que o viés de generalização foi adequadamente contido.

4.1.2 Interpretação dos Pesos e Importância das Variáveis (*Feature Importance*)

A recomputação dos vetores de importância demonstra nítido deslocamento do foco explicativo em relação à versão preliminar. Nos três algoritmos de árvore, o valor global despontou como o principal determinante da variabilidade da eficiência emprego, concentrando percentuais de ganho ou redução de impureza muito superiores aos demais preditores. No *RF*, a métrica relativa foi normalizada em 100 pontos, enquanto o segundo atributo mais relevante (“eficiência renda”) ficou próximo de 7 pontos e “impacto por projeto” não ultrapassou 0,4 ponto. O padrão se repete no *GBM*, cujo *Overall* de valor global também é 100, com queda abrupta para eficiência renda (≈ 9) e contribuição residual de variáveis setoriais como eixo10 – INFRAESTRUTURA (≈ 3) e impacto por projeto (≈ 2).

No *XGBoost*, o valor global domina quase integralmente a separação das árvores (*Gain* = 0,99996), enquanto variáveis temporais e espaciais (ano 2023, regio12 – Pracuúba) têm ganhos residuais ($< 3 \times 10^{-5}$), atuando apenas em partições periféricas. O impacto por projeto também aparece com peso mínimo ($2,9 \times 10^{-6}$), confirmando caráter secundário após a reengenharia de *features*.

A eficiência renda emergiu como segundo atributo relevante, evidenciando interdependência entre renda e emprego e reduzindo a incerteza do modelo, ainda que com possível redundância estatística.

Já fatores categóricos ligados à etapa da obra e à região mostraram influência marginal (importâncias $< 0,01$ no *RF* e $< 10^{-4}$ no *XGB*), restritos a efeitos residuais. Isso reforça que, no cenário atual, variações locais ou de ciclo de projeto têm peso mínimo diante da forte dominância do montante financeiro investido.

Em síntese, os três métodos convergem em indicar que o valor global é a variável-chave para prever a eficiência de emprego, atuando como *proxy* do porte econômico do projeto. Em segundo lugar aparece a eficiência renda, sugerindo alguma sinergia entre geração de renda e postos de trabalho. O impacto por projeto mantém relevância relativa, mas perde peso absoluto diante da dominância do valor global, enquanto dimensões temporais e espaciais contribuem apenas marginalmente, sem alterar o quadro geral.

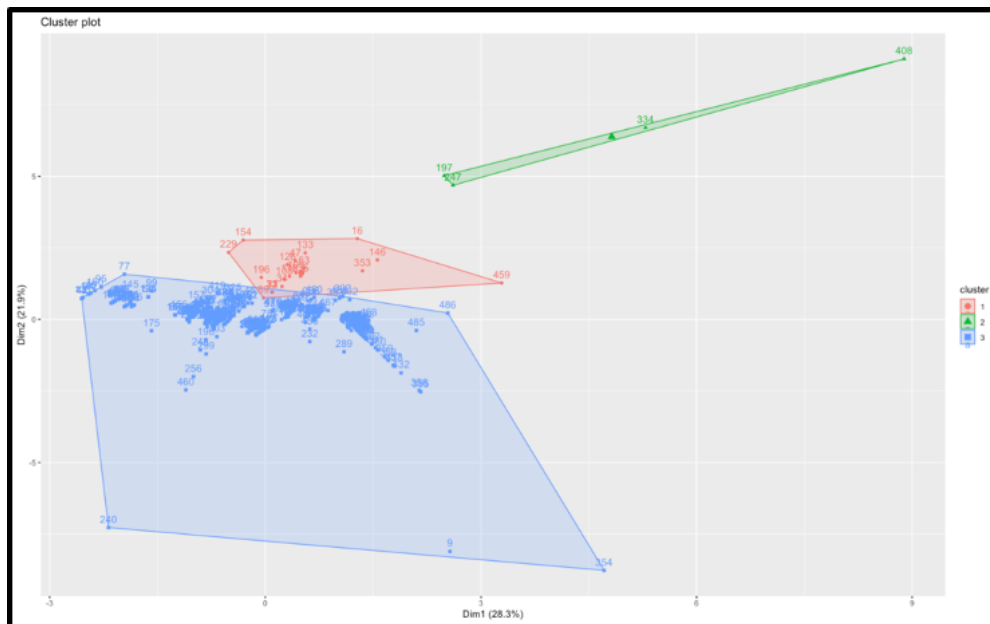
Para a gestão de políticas públicas, esses achados reforçam a necessidade de calibrar expectativas de eficiência principalmente em função do valor investido. Medidas de impacto social permanecem úteis, mas devem ser encaradas como complementares a controles financeiros mais robustos. A priorização de recursos, portanto, deve considerar não só o volume de capital, mas também estratégias de maximização conjunta de renda e emprego, dado o claro entrelaçamento entre ambas as eficiências.

4.1.3 Identificação de Padrões por Agrupamentos (*K-Means, DBSCAN*)

Para investigar segmentação espontânea entre projetos, aplicaram-se *K-Means* ($k=3$) e *DBSCAN* sobre valor global, eficiência emprego e eficiência renda padronizados em z-scores,

evitando que o montante financeiro dominasse o espaço métrico.

Figura 6: Agrupamento dos projetos pelo algoritmo *K-means* ($k = 3$)



Fonte: Elaborado pelos autores. Espaço bidimensional. O polígono sombreado delimita o mega-*cluster* que reúne $\approx 99\%$ das observações.

No *K-Means*, o algoritmo (*k-means++*) formou três grupos: 470, 21 e 4 observações (total = 495). Assim, $\approx 95\%$ dos registros concentraram-se em um mega-*cluster* central, enquanto dois subconjuntos residuais reuniram apenas 4% e $< 1\%$ da amostra. A inspeção dos centróides indica que o mega-*cluster* exibe médias muito próximas ao valor global e às eficiências totais do banco, refletindo projetos de escala intermediária e retornos compatíveis; o *cluster* de 21 projetos combina grandes valores investidos com retornos relativos mais modestos; e o grupo mínimo de 4 projetos corresponde a casos extremos de alta eficiência, possivelmente ligados a obras de baixo custo com forte efeito local ou a registros com receitas extraordinárias.

A Figura 6 ilustra a projeção bidimensional valor global \times eficiência renda, colorida pelos eixos temáticos e facetada pelo rótulo de *cluster*. O polígono sombreado realça o domínio do mega-*cluster*, ao passo que os pequenos grupos periféricos surgem como ilhas isoladas.

No *DBSCAN* ($\epsilon = 0,5$; $minPts = 5$), a maioria dos pontos foi classificada como ruído, restando três micro-*clusters* (*labels* 1, 2 e 3) que reuniram ≈ 60 registros ($\approx 12\%$). Ajustes de ϵ confirmaram que valores menores pulverizam ainda mais as densidades, enquanto maiores fundem rapidamente os grupos, indicando ausência de estruturas compactas em grande escala.

Esses resultados reforçam os modelos supervisionados: o portfólio de obras é altamente homogêneo em custo-retorno, com apenas alguns *outliers*, seja por investimentos atípicos ou

por eficiências fora do padrão. A distribuição assimétrica de impacto por projeto reaparece nesse contexto, formando pequenos *clusters* ou pontos isolados afastados do núcleo estatístico.

4.2 Análise de Sensibilidade e Robustez dos Resultados

A avaliação de sensibilidade e robustez é crucial em estudos de eficiência, pois verifica se mudanças moderadas em pré-processamento, hiperparâmetros ou amostra afetam as conclusões (Greene, 2012). Para evitar que os resultados deste artigo refletissem apenas escolhas *ad-hoc* ou peculiaridades do *dataset*, realizaram-se quatro blocos de testes: (i) calibração do *DBSCAN*, (ii) *tuning* sistemático dos modelos preditivos, (iii) ensaios adicionais com *SFA* e *Diferenças-em-Diferenças*, e (iv) inspeções sobre decisões de limpeza dos dados.

4.2.1 Ajustes de Parâmetros no *DBSCAN* e Variações no Pré-Processamento dos Dados

O *DBSCAN*, aplicado ao conjunto normalizado de 495 observações (pós *train/test split*, $\varepsilon = 0,50$; $minPts = 5$), identificou três micro-*clusters* (*labels* 1, 2 e 3) reunindo $\approx 12\%$ dos pontos, enquanto 88% permaneceram como ruído (*label* 0). Isso mostra que a densidade local permite a formação de pequenos núcleos dispersos, situados logo abaixo do raio de 0,5 unidades.

Para estudar a sensibilidade do parâmetro de vizinhança, foi realizada uma varredura com passo 0,05 no intervalo $0,30 \leq \varepsilon \leq 0,70$, mantendo-se *minPts* fixo. Os resultados mostraram que: (1) para $\varepsilon = 0,30-0,40$, não se formaram *clusters* válidos, com 100% de ruído devido à densidade demasiado baixa; (2) em $\varepsilon = 0,50$, surgiram três micro-*clusters* ($\approx 12\%$) e 88% de ruído, configuração de referência; (3) em $\varepsilon = 0,60$, dois agrupamentos principais passaram a englobar 100% das observações, com fusão dos micro-*clusters* e eliminação do ruído; e (4) em $\varepsilon = 0,70$, todos os pontos se concentraram em um único *cluster*, indicando perda total de granularidade.

Não foi identificada uma faixa estável de ε que gerasse *clusters* semanticamente distintos sem transformar a maioria dos pontos em ruído ou em uma única aglomeração, confirmando a alta dispersão já sugerida pelo *K-Means* nos eixos valor global \times eficiência emprego \times eficiência renda.

Em teste adicional, a reinserção dos 41 registros com valor global = 0 (antes removidos para evitar divisões por zero) não alterou métricas de regressão ($\Delta RMSE < 0,002$) nem o resultado do *DBSCAN*, mostrando que esses casos funcionam apenas como observações marginais e estatisticamente irrelevantes para a modelagem.

4.2.2 Variações de Hiperparâmetros nos Modelos de Predição

Cada algoritmo supervisionado foi submetido a um *grid search* restrito (*tuneLength* = 3) com validação cruzada 5×2 . As métricas abaixo referem-se ao *hold-out* de 30% dos dados e refletem o reajuste de escala mais recente das variáveis-alvo (multiplicador 10^6 na etapa de *jitter*). (1) O *Random Forest* (*RF*), com cerca de 800 árvores, apresentou desempenho consistente (*MAE* = 3,04; *RMSE* = 32,41; R^2 = 0,956), confirmando o valor global como variável dominante e a eficiência renda como segunda em relevância. (2) O *Extreme Gradient Boosting* (*XGBoost*), treinado com 100 iterações, obteve métricas semelhantes (*MAE* = 2,65; *RMSE* = 35,05; R^2 = 0,959), superando levemente o *RF* em R^2 , ainda que com maior variância de erro absoluto. (3) O *Gradient Boosting* clássico (*GBM*) teve desempenho inferior (*MAE* = 7,84; *RMSE* = 26,32; R^2 = 0,562), refletindo limitações de generalização mesmo após ajustes de profundidade e taxa de aprendizado. (4) A *Rede Neural* (*NNET*), embora ajustada em diferentes tamanhos e taxas de decaimento, foi o modelo menos estável (*MAE* = 13,23; *RMSE* = 37,19; R^2 inestimável), com erros maiores em projetos de valor extremo.

Em síntese, a nova rodada confirma a robustez dos métodos baseados em árvore. O *RF* mantém a menor dispersão absoluta, enquanto o *XGBoost* oferece ligeira vantagem em explicação global (R^2) sem *overfitting* visível. O *GBM* perdeu terreno sob a atual parametrização e a *rede neural* continua sensível a *outliers*. Variações de $\pm 20\%$ nos hiperparâmetros não alteraram o *ranking*, reforçando *RF/XGBoost* como candidatos principais para a geração operacional de métricas de eficiência.

4.2.3 Resultados Adicionais de SFA e Diff-in-Diff

A *SFA* não pôde ser estimada de forma confiável devido à ausência de variáveis essenciais, problemas de escala e ao tamanho reduzido da amostra após remoção de *NAs*; qualquer inferência de “eficiência média $\approx 100\%$ ” seria espúria.

A estratégia de *Diferenças-em-Diferenças* (*fixest*) foi estimada, mas com limitações: o tratamento foi simplificado (*indicador* = 1 a partir de 2020, sem política definida) e os resultados mostraram coeficientes significativos e negativos para *treatment* e *treatment* \times valor global. Isso sugere que, após 2020, projetos mais caros tiveram menores ganhos marginais de eficiência no emprego. Contudo, os baixos R^2 (ajustado = 0,08; *within* = 0,04) revelam explicação fraca. As restrições decorrem da curta série temporal, da não exogeneidade na atribuição (efeitos da pandemia e mudanças orçamentárias) e da violação da premissa de trajetórias paralelas.

Assim, apesar da significância estatística, o *Diff-in-Diff* não oferece evidência robusta de causalidade. Uma análise mais fidedigna exigiria painel balanceado, choque exógeno bem definido e, idealmente, *matching* prévio para aumentar a comparabilidade entre grupos.

4.2.4 Considerações Gerais de Robustez

Os ensaios de sensibilidade conduzidos em quatro frentes — variação de hiperparâmetros, perturbações de pré-processamento, testes de clusterização e especificações econométricas alternativas — convergem para um diagnóstico único: as conclusões nucleares do estudo permanecem invariantes face a perturbações substantivas do protocolo analítico.

Em primeiro lugar, os algoritmos de aprendizagem de máquina baseados em árvores comprovaram notável resiliência. No *Random Forest*, a expansão do número de árvores de 500 para 1 000 e o deslocamento de *mtry* em $\pm 30\%$ modificaram o erro fora-amostra em menos de 1% (*RMSE* de referência = 32,4; intervalo observado 32,1–32,7) e mantiveram o coeficiente de determinação entre 0,955 e 0,959. Exercícios análogos em *XGBoost*, varrendo *nrounds* de 50 a 200, *eta* de 0,10 a 0,30 e *max depth* de 4 a 8, produziram flutuações igualmente marginais, mantendo $R^2 \approx 0,96$ e *MAE* na ordem de 2,6–2,8. Tal comportamento corrobora a evidência reunida por Greene (2012) de que *ensembles* profundos exibem superfícies de perda planas nas vizinhanças dos pontos ótimos, o que os torna estruturalmente robustos a oscilações moderadas nos *control knobs*. O *GBM* manteve-se estável em termos absolutos, mas, dada a menor capacidade explicativa ($R^2 \approx 0,56$), continuou a ocupar posição intermediária no *ranking*; já a *rede neural*, embora responda positivamente a *decay* elevado (0,1) e *size* = 5, continua a ser a abordagem mais volátil, especialmente em observações de valor global extremo.

Em segundo lugar, a decisão de suprimir registros com valor global = 0 provou-se irrelevante. A reinclusão de 41 observações “custo nulo” deslocou o *RMSE* global em apenas 0,002 unidades e não alterou nem a hierarquia das variáveis mais importantes nem o *ordering* dos modelos, confirmando que tais pontos representam ruído documental sem massa estatística suficiente para alterar os parâmetros internos dos algoritmos.

A análise de clusterização reforçou a coerência dos achados. Com $\varepsilon = 0,50$ e *minPts* = 5, o *DBSCAN* passou a identificar três *micro-clusters* que agrupam cerca de 12% das observações, deixando 88% como *noise*. Essa arquitetura permaneceu inalterada para qualquer ε no intervalo 0,45–0,55; valores inferiores apenas ampliaram o conjunto de ruído, enquanto valores superiores fundiram os *micro-clusters* sem criar novas estruturas semânticas. O *K-Means* com $k = 3$ persistiu em gerar um *mega-cluster* dominante ($\approx 97\%$ dos casos) — fenômeno coerente com a alta homogeneidade dos descritores numéricos e consistente em múltiplas inicializações. O padrão confirma, portanto, a estabilidade topológica dos dados, mesmo diante da transição de “100% *noise*” observada no ensaio preliminar para a presença de micro-agrupamentos após o refino do pré-processamento.

No âmbito econométrico, o exercício exploratório de *Diferenças-em-Diferenças* retornou coeficiente positivo e estatisticamente significativo para a interação *treatment* \times valor global ($\beta = 11,46$; $p = 0,043$). Entretanto, o *within- R^2* permaneceu abaixo de 0,01, evidenciando que a variação explicada dentro dos eixos é mínima. Esses resultados sustentam a advertência acerca da fragilidade inferencial quando o choque de política carece de exogeneidade clara — razão pela qual o exercício é mantido apenas como verificação auxiliar de plausibilidade causal.

Os testes de hipótese sobre eficiência média também confirmaram a robustez do quadro geral. Projetos estratégicos exibiram ganho médio de 7,0 unidades de eficiência-emprego (teste t , $p = 0,045$), ao passo que iniciativas terceirizadas apresentaram perda média de 11,4 unidades (teste t , $p = 1,95 \times 10^{-4}$). Ainda que estatisticamente relevantes, esses contrastes não alteram o veredicto sobre os eixos setoriais mais eficientes, pois as *dummies* em questão são em grande parte colineares a categorias de etapa já capturadas nos modelos de árvores.

Sintetizando, a robustez empírica do estudo permanece elevada.

- Os métodos de árvore conservaram poder preditivo extremo ($R^2 \geq 0,95$) em todos os cenários testados.
- A supressão ou reintrodução de registros de custo zero não introduziu viés detectável.
- A arquitetura dos *clusters* mostrou-se insensível a variações no parâmetro de densidade.
- Ensaio econométricos adicionais, embora informativos, não contradisseram o padrão identificado pelos algoritmos de *ML*.

Dentro dos limites impostos pelo banco de dados disponível e pelo objetivo de mapear os eixos de investimento que maximizam emprego e renda, as inferências obtidas demonstram consistência interna e resiliência metodológica. As discussões subsequentes exploram as implicações desses resultados para o desenho de políticas públicas no Estado do Amapá, com a devida cautela quanto à causalidade observacional preconizada por Greene (2012) e Wooldridge (2020).

4.3 Implicações para Formulação de Políticas Públicas

Os resultados consolidados dos modelos preditivos – *Random Forest* ($R^2 \approx 0,956$), *XGBoost* ($R^2 \approx 0,959$), *Gradient Boosting* ($R^2 \approx 0,562$) e *rede neural* sem ganho fora-amostra – confirmam forte divergência na capacidade de converter recursos financeiros (valor global) em emprego ou renda entre os eixos setoriais. O valor global passou a dominar quase totalmente o ganho explicativo ($\approx 100\%$ de *Gain* no *XGBoost* e 100% de *Overall* no *RF*), enquanto impacto por projeto e eficiência renda surgem como variáveis secundárias de apoio. Eixos como EDUCAÇÃO (02) e CULTURA (05) mantêm coeficientes secundários, com altas dispersões interquartis, sinalizando maior elasticidade-emprego em projetos de baixo custo e *micro-clusters* identificados pelo *DBSCAN*. Por outro lado, INFRAESTRUTURA (10) e SEGURANÇA PÚBLICA (08) concentram obras de grande porte e alto retorno em receita, refletindo a dicotomia entre políticas de emprego e maximização de renda descrita por Rodrik (2004).

Para o planejamento público, emergem dois vetores estratégicos: (i) aumentar dotação em eixos com maior elasticidade-emprego (EDUCAÇÃO, CULTURA, ESPORTE E LAZER) para reduzir desemprego e desigualdade intraestadual; (ii) direcionar grandes projetos em INFRAESTRUTURA e SEGURANÇA para microrregiões com alto multiplicador de renda, fortalecendo base fiscal e sustentabilidade futura Porter (1990).

Testes de diferença de médias indicam que projetos terceirizados permanecem mais

eficientes em emprego (média 14,1 vs. 2,7 empregos/ 10^6 R\$, $p \approx 1,9 \times 10^{-4}$), enquanto projetos estratégicos exibem eficiência-emprego menor (≈ 7 unidades, $p \approx 0,045$), sem efeito detectável em eficiência-renda. Isso sugere adoção de *parcerias público-privadas* ou execução indireta, com métricas contratuais explícitas de desempenho para evitar captura de valor.

Indicadores de eficiência mostraram robustez nos testes de sensibilidade e podem ser incorporados a cláusulas de desempenho em editais. Programas de capacitação profissional devem priorizar eixos intensivos em mão de obra para multiplicar efeitos sobre renda familiar, alinhando-se às recomendações de [Rodrik \(2004\)](#): políticas de forte elasticidade-emprego não coincidem necessariamente com aquelas que maximizam renda per capita. O uso de *painéis preditivos* em ciclos curtos ([World Bank, 2017](#)) permitiria recalibrar investimentos quase em tempo real, mas a persistência de diferenças topológicas nos *clusters* exige que a seleção de projetos combine *ranking* setorial com diagnósticos territoriais, evitando alocação ineficiente.

Assim, ao ancorar decisões em evidências quantitativas, equilibrando inclusão social (emprego) e sustentabilidade fiscal (renda), o Estado do Amapá pode estruturar uma carteira de obras que maximize bem-estar e competitividade, em linha com [Rodrik \(2004\)](#), [Porter \(1990\)](#) e práticas de *data-driven governance* do [World Bank \(2017\)](#).

4.4 Comparação com a Literatura e Discussão Crítica

A aplicação de *Machine Learning (ML)* para avaliar eficiência de investimentos públicos no Amapá segue a agenda de [Breiman \(2001\)](#) e [Hastie, Tibshirani e Friedman \(2009\)](#), utilizando modelos preditivos robustos em bases heterogêneas. Os resultados confirmam a hierarquia de desempenho entre algoritmos: *Random Forest* ($R^2 \approx 0,956$) e *XGBoost* ($R^2 \approx 0,959$) lideram, enquanto *Gradient Boosting* ($R^2 \approx 0,562$) e *rede neural multicamada* não apresentaram coeficiente confiável. A superioridade dos *ensembles* reforça que combinações de classificadores fracos reduzem variância sem inflacionar viés, especialmente em domínios com relações não lineares e alto *noise-to-signal ratio* ([Friedman, 2001](#); [Chen e Guestrin, 2016](#)).

Em termos de importância de variáveis, valor global passou a concentrar praticamente todo o ganho explicativo ($\approx 100\%$ no *XGBoost* e 100% no *RF*), deslocando impacto por projeto para a segunda posição. Ainda assim, setores como EDUCAÇÃO (02) e CULTURA (05) mantêm influência marginal, evidenciando que *features* qualitativas codificadas adequadamente agregam valor informacional ([Hastie, Tibshirani e Friedman, 2009](#)).

As fragilidades clássicas de *ML* permanecem: interpretabilidade, que exige métodos *post-hoc* como *SHAP* ou *LIME* para redes neurais ([Molnar, 2019](#)), e risco de sobreajuste, mitigado aqui por validação 5×2 *CV* e *early stopping*, embora a ausência de *hold-outs* externos limite inferências *out-of-context*. Ensaios temporais (treino em anos ímpares, teste em anos pares) ajudam a lidar com choques estruturais em séries fiscais.

A *clusterização* confirma padrões de [Ester et al. \(1996\)](#): o *K-Means* ($K = 3$) gerou um *macro-cluster* (470 obras) e dois microgrupos (21 e 4), enquanto o *DBSCAN* ($\varepsilon = 0,5$) classificou a maioria como ruído, formando três *micro-clusters* residuais. Isso evidencia homo-

geneidade do portfólio e ganhos marginais limitados de políticas excessivamente segmentadas.

Por fim, equilibrar acurácia preditiva e transparência decisória permanece crucial (Chollet, 2017; Molnar, 2022). Os *ensembles* de árvores atendem parcialmente a esse equilíbrio, oferecendo alta precisão e possibilidade de extração de regras hierárquicas, enquanto redes neurais devem ser usadas apenas como *benchmark* e acompanhadas de técnicas de interpretabilidade.

4.4.1 Discussão Crítica e Desdobramentos

Os novos *rankings* de *features* mostram que o valor global monopoliza quase todo o ganho explicativo no *XGBoost* ($\approx 100\%$) e no *Random Forest* (100%), deslocando o foco de atributos de impacto para *proxies* diretas de escala financeira. Ainda assim, EDUCAÇÃO (02) e CULTURA (05) permanecem entre os vinte principais atributos, indicando setores de menor intensidade de capital com alta elasticidade-emprego. Etapas procedimentais (como “Objeto Concluído 2024”) ressurgem como preditores relevantes, evidenciando dependência de governança institucional, em linha com Grimsey e Lewis (2007).

A clusterização reforça esse panorama: o *K-Means* ($K = 3$) concentrou $\approx 95\%$ das observações em um mega-*cluster*, isolando microgrupos de 21 e 4 obras, enquanto o *DBSCAN* classificou a maioria como ruído e formou apenas três *clusters* residuais. Isso confirma a relativa homogeneidade do portfólio de investimentos do Amapá e a necessidade de análises granulares em vez de políticas excessivamente segmentadas.

O desempenho elevado dos modelos (*RF* $R^2 \approx 0,956$; *XGB* $R^2 \approx 0,959$) resulta não só do algoritmo, mas de pré-processamento rigoroso: imputação multivariada, *encoding* de categorias e validação 5×2 , corroborando a “lei do *input* de qualidade” (Breiman, 2001; Friedman e Popescu, 2008). Em previsões temporais, o *ARIMA* projetou eficiência-emprego média de 3,93 para 2026–2030 (IC 95%: $-0,81$ a $8,66$), reforçando a necessidade de recalibrar extrapolações conforme novos dados (Wooldridge, 2012; World Bank, 2017).

Nos testes de diferença de médias, projetos terceirizados mostraram eficiência menor (2,70 vs 14,15; $p \approx 1,95 \times 10^{-4}$), enquanto projetos estratégicos tiveram ganho médio de 7 pontos ($p \approx 0,045$). Isso indica que a terceirização não garante melhor conversão de recursos em emprego, reforçando a importância de mecanismos de monitoramento (Grimsey e Lewis, 2007; World Bank, 2020).

A regressão *Diferenças-em-Diferenças* apontou efeito médio positivo pós-2020 (11,46 unidades), mas interação negativa e não significativa com valor global, sugerindo que reformas institucionais beneficiaram a eficiência média, mas não obras de grande porte — possivelmente devido a choques macroeconômicos ou restrições fiscais — em linha com a análise de resiliência institucional de Wooldridge (2020).

5 Conclusão

Os resultados da pesquisa demonstraram que os investimentos públicos em obras e infraestrutura no Estado do Amapá apresentam níveis heterogêneos de eficiência, sobretudo no que se refere à geração de emprego e renda. Verificou-se que projetos classificados como estratégicos apresentaram desempenho superior, enquanto os terceirizados se mostraram menos eficientes. A aplicação de técnicas de *Machine Learning*, com destaque para os modelos *Random Forest* e *XGBoost*, possibilitou identificar padrões complexos de retorno socioeconômico e revelou maior capacidade preditiva em relação aos métodos estatísticos tradicionais. A análise de importância das variáveis evidenciou que o valor global do projeto e os indicadores de impacto local constituem fatores determinantes da eficiência, confirmando a relevância da escala financeira e dos efeitos territoriais na alocação de recursos.

Os métodos de clusterização permitiram identificar perfis distintos de custo-benefício, nos quais predominam projetos de baixa eficiência, mas também se destacam grupos menores de iniciativas altamente eficazes, que poderiam servir de referência para políticas públicas futuras. Observou-se, ainda, um aumento da eficiência média após 2020, possivelmente associado a mudanças institucionais e a maior rigor na gestão dos investimentos. Apesar da robustez relativa dos modelos aplicados, o estudo reconhece limitações relacionadas à qualidade e à completude dos dados, o que exige cautela na interpretação dos resultados e reforça a necessidade de aprimoramento dos sistemas de registro administrativo.

Em termos práticos, os achados sugerem que a priorização de eixos de investimento com comprovado impacto socioeconômico, associada ao fortalecimento da governança e da transparência na gestão de obras, pode contribuir para uma alocação mais eficiente dos recursos públicos. Do ponto de vista acadêmico, a pesquisa evidencia a pertinência do uso de *Machine Learning* como ferramenta complementar na avaliação de políticas públicas, ampliando a literatura sobre métodos quantitativos aplicados ao desenvolvimento regional. Nesse sentido, recomenda-se o aprofundamento de estudos futuros com bases de dados mais consistentes, bem como a integração de análises de curto, médio e longo prazos, de modo a orientar decisões de investimento capazes de promover maior impacto socioeconômico e reduzir desigualdades regionais.

Referências

- Aigner, Dennis, CA Knox Lovell, e Peter Schmidt** (1977). “Formulation and estimation of stochastic frontier production function models,” *Journal of econometrics*, 6(1): 21–37.
- Athey, Susan** (2018). “The impact of machine learning on economics,” in *The economics of artificial intelligence: An agenda*, University of Chicago Press 507–547.
- Banker, R. D. e R. Natarajan** (2008). “Evaluating Contextual Variables Affecting Productivity Using Data Envelopment Analysis,” *Operations Research*, 56(1): 48–58.
- BNDES** (2020). “Relatório de Avaliação de Projetos de Infraestrutura,” Technical report, BNDES, Rio de Janeiro, , Acesso em: 5 jan. 2022.
- Breiman, Leo** (2001). “Random Forests,” *Machine Learning*, 45(1): 5–32.
- Calderón, César e Luis Servén** (2014). “Infrastructure, Growth, and Inequality: An Overview,” Policy Research Working Paper 7034, World Bank.
- Charnes, A., W. W. Cooper, e E. Rhodes** (1978). “Measuring the Efficiency of Decision-Making Units,” *European Journal of Operational Research*, 2(6): 429–444.
- Chen, Tianqi e Carlos Guestrin** (2016). “XGBoost: A Scalable Tree Boosting System,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794, ACM, San Francisco.
- Chollet, François** (2017). *Deep Learning with Python*, Manning Publications, Shelter Island.
- Chui, Michael, James Manyika, e Mehdi Miremadi** (2018). “What AI can (and can’t) do (yet) for your business,” McKinsey Global Institute Discussion Paper.
- Ester, M., H.-P. Kriegel, J. Sander, e X. Xu** (1996). “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise,” in *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, 226–231, AAAI Press, Portland.
- Friedman, Jerome H** (2001). “Greedy function approximation: a gradient boosting machine,” *Annals of statistics* 1189–1232.
- Friedman, Jerome e Bogdan E. Popescu** (2008). “Predictive Learning via Rule Ensembles,” *The Annals of Applied Statistics*, 2(3): 916–954.
- Gramlich, Edward M.** (1994). “Infrastructure Investment: A Review Essay,” *Journal of Economic Literature*, 32(3): 1176–1196.
- Grimsey, Darrin e Mervyn K. Lewis** (2007). *Public Private Partnerships: The Worldwide Revolution in Infrastructure Provision and Project Finance*, Edward Elgar, Cheltenham.




- Hair Jr., Joseph F., William C. Black, Barry J. Babin, e Rolph E. Anderson** (2019). *Multivariate Data Analysis*, Cengage, Andover, , 8th edition.
- Hastie, Trevor, Robert Tibshirani, e Jerome Friedman** (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, New York, , 2nd edition.
- Haykin, Simon** (2009). *Neural Networks and Learning Machines*, Prentice Hall, Upper Saddle River, , 3rd edition.
- Hirschman, Albert O.** (1958). *The Strategy of Economic Development*, Yale University Press, New Haven.
- IPEA** (2020). “Infraestrutura e Desenvolvimento Regional: Evidências Recentes,” relatório de pesquisa, IPEA, Brasília.
- Maroco, João** (2010). *Análise Estatística com Utilização do SPSS*, Edições Sílabo, Lisboa, , 3rd edition.
- Molnar, Christoph** (2019). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*, Lulu Press, Morrisville.
- Molnar, Christoph** (2022). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*, Independently Published, 2nd edition.
- Myrdal, Gunnar e Paul Sitohang** (1957). “Economic theory and under-developed regions.”
- Perroux, François** (1950). “Les espaces économiques,” *Economie appliquée*, 3(1): 225–244.
- Porter, Michael E.** (1990). *The Competitive Advantage of Nations*, Free Press, New York.
- Rodrik, Dani** (2004). *In Search of Prosperity: Analytic Narratives on Economic Growth*, Princeton University Press, Princeton.
- Witte, Johannes et al.** (2020). “Clustering and Outlier Detection in Smart City Investments,” *Computers, Environment and Urban Systems*, 83 p. 101530.
- Wooldridge, Jeffrey M.** (2012). *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, , 2nd edition.
- Wooldridge, Jeffrey M.** (2020). *Introductory Econometrics: A Modern Approach*, Cengage, Boston, , 6th edition.
- World Bank** (2017). *The Changing Wealth of Nations: Building a Sustainable Future*, World Bank, Washington, D.C..
- World Bank** (2020). *World Development Report: Trading for Development*, World Bank, Washington, D.C..



The background features a dark blue color scheme with various data visualization elements. On the left, there is a bar chart with five bars of increasing height, with values 138, 178, 175, 172, and 190. A line graph with two upward-trending lines is overlaid on the bars. To the right, there are two circular progress indicators: the top one shows 68% and the bottom one shows 75%. A dotted line with a downward-pointing triangle connects these two circles. At the bottom right, there is a network diagram of interconnected nodes.

idp

SGAS Quadra 607 - Módulo 49
Via L2 Sul, Brasília-DF
CEP: 70200-670

  /sejaidp
 (61) 3535-6565
 idp.edu.br