Revista Brasileira de Informática na Educação – RBIE
Brazilian Journal of Computers in Education
(ISSN online: 2317-6121; print: 1414-5685)
https://sol.sbc.org.br/journals/index.php/rbie

Submission: 01/12/2025; 1st round notif.: 11/02/2026; New version: 06/03/2026; 2nd round notif.: 22/04/2026
Camera ready: 04/05/2026; Edition review: 19/05/2026; Available online: 19/05/2026; Published: 19/05/2026

Aplicação de Técnicas de Machine Learning e Deep Learning para Prever o IDEB Municipal Piauiense nos Anos Iniciais do Ensino Fundamental

Application of Machine Learning and Deep Learning Techniques to Predict the Piauiense Municipal IDEB in the Initial Years of Elementary School

Aplicación de técnicas de aprendizaje automático y aprendizaje profundo para predecir el IDEB municipal piauiense en los años iniciales de la educación primaria

Maria Eva Clemencia Fonseca de Castro Silva
Universidade Federal do Piauí
ORCID: 0009-0000-8230-0274
clemencia182017@gmail.com

Ivan Saraiva Silva
Universidade Federal do Piauí
ORCID: 0000-0002-5705-6932
ivan@ufpi.edu.br

Vinícius Ponte Machado
Universidade Federal do Piauí
ORCID: 0000-0003-3391-8443
vinicius@ufpi.edu.br

Resumo

Este estudo investiga a aplicação de modelos preditivos para estimar o Índice de Desenvolvimento da Educação Básica (IDEB) nos anos iniciais do ensino fundamental nos municípios do estado do Piauí, a partir de variáveis educacionais, socioeconômicas e financeiras. A pesquisa utilizou dados públicos e, após rigoroso processo de limpeza, agregação, imputação e seleção de atributos, constituiu uma base multivariada final composta por 1.262 registros e 126 variáveis explicativas, extraídas exclusivamente de fontes oficiais. Foram avaliados dez algoritmos de Machine Learning (ML), incluindo regressões lineares e penalizadas, árvores de decisão, métodos baseados em vizinhança, ensembles e support vector regression e sete arquiteturas de Deep Learning (DL), abrangendo redes do tipo Multilayer Perceptron (MLP), variantes com Dropout e Batch Normalization, arquiteturas convolucionais (CNN), recorrentes (LSTM) e híbridas. Os experimentos seguiram uma abordagem quantitativa, com divisão dos dados em conjuntos de treino e teste (70/30), validação cruzada k-fold, normalização quando necessário e análise de importância de variáveis. O algoritmo Extreme Gradient Boosting (XGBoost) apresentou o melhor desempenho médio entre os modelos avaliados, alcançando $R^2 = 0{,}5542$ *no conjunto de teste e* $R^2 = 0{,}5455 \pm 0{,}0293$ *em validação cruzada, com menores erros médios e maior estabilidade relativa em comparação às abordagens lineares e às redes neurais profundas. Além disso, a análise de importância dos preditores identificou o PIB per capita municipal, a taxa de distorção idade-série e a proporção de docentes sem formação superior como os fatores mais relevantes para explicar a variação do IDEB nos municípios piauienses. Os achados contribuem para o aprimoramento de diagnósticos regionais e para a proposição de uma abordagem replicável de apoio ao monitoramento educacional em escala municipal.*

Palavras-Chave: *Educação Básica; IDEB; Modelagem Preditiva; Machine Learning; Deep Learning; Ensembles Baseados em Árvores; Políticas Educacionais Baseadas em Evidências.*

Abstract

This study investigates the application of predictive models to estimate the Basic Education Development Index (IDEB) in the early years of elementary school in municipalities in the state of Piauí, based on educational, socioeconomic, and financial variables. The research used public data and, after a rigorous process of cleaning, aggregation, imputation, and attribute selection, constituted a final multivariate database composed of 1,262 records and 126 explanatory variables, extracted exclusively from official sources. Ten Machine Learning (ML) algorithms were evaluated, including linear and penalized regressions, decision trees, neighborhood-based methods, ensembles, and support vector regression, and seven Deep Learning (DL) architectures, encompassing Multilayer Perceptron

Cite as: Silva, M. E. C. F. C., Silva, I. S., & Machado, V. P. (Year). Aplicação de Técnicas de Machine Learning e Deep Learning para Prever o IDEB Municipal Piauiense nos Anos Iniciais do Ensino Fundamental. Revista Brasileira de Informática na Educação, vol. 34, pp. 717-738. https://doi.org/10.5753/rbie.2026.7071

(MLP) networks, variants with Dropout and Batch Normalization, convolutional (CNN), recurrent (LSTM), and hybrid architectures. The experiments followed a quantitative approach, with data splitting into training and test sets (70/30), k-fold cross-validation, normalization when necessary, and variable importance analysis. The Extreme Gradient Boosting (XGBoost) algorithm presented the best average performance among the evaluated models, reaching $R^2 = 0.5542$ *in the test set and* $R^2 = 0.5455 \pm 0.0293$ *in cross-validation, with lower average errors and greater relative stability compared to linear approaches and deep neural networks. Furthermore, the importance analysis of the predictors identified municipal GDP per capita, age-grade distortion rate, and the proportion of teachers without higher education as the most relevant factors to explain the variation in IDEB (Basic Education Development Index) in the municipalities of Piauí. The findings contribute to the improvement of regional diagnoses and to the proposal of a replicable approach to support educational monitoring at the municipal level.*
Keywords: *Basic Education; IDEB; Predictive Modeling; Machine Learning; Deep Learning; Tree-Based Ensembles; Evidence-Based Educational Policies.*

Resumen

Este estudio investiga la aplicación de modelos predictivos para estimar el Índice de Desarrollo de la Educación Básica (IDEB) en los primeros años de la escuela primaria en municipios del estado de Piauí, con base en variables educativas, socioeconómicas y financieras. La investigación utilizó datos públicos y, tras un riguroso proceso de limpieza, agregación, imputación y selección de atributos, se conformó una base de datos multivariante final compuesta por 1262 registros y 126 variables explicativas, extraídas exclusivamente de fuentes oficiales. Se evaluaron diez algoritmos de aprendizaje automático (AA), incluyendo regresiones lineales y penalizadas, árboles de decisión, métodos basados en vecindad, ensambles y regresión de vectores de soporte, y siete arquitecturas de aprendizaje profundo (AP), que abarcan redes perceptrón multicapa (MLP), variantes con dropout y normalización por lotes, redes neuronales convolucionales (CNN), redes recurrentes (LSTM) y arquitecturas híbridas. Los experimentos siguieron un enfoque cuantitativo, con división de datos en conjuntos de entrenamiento y prueba (70/30), validación cruzada k-fold, normalización cuando fue necesario y análisis de importancia de las variables. El algoritmo Extreme Gradient Boosting (XGBoost) presentó el mejor rendimiento promedio entre los modelos evaluados, logrando $R^2 = 0.5542$ *en el conjunto de prueba y* $R^2 = 0.5455 \pm 0.0293$ *en la validación cruzada, con errores promedio más bajos y mayor estabilidad relativa en comparación con los enfoques lineales y las redes neuronales profundas. Además, el análisis de importancia de los predictores identificó el PIB municipal per cápita, la tasa de distorsión edad-grado y la proporción de docentes sin educación superior como los factores más relevantes para explicar la variación en el IDEB (Índice de Desarrollo de la Educación Básica) en los municipios de Piauí. Los resultados contribuyen a la mejora de los diagnósticos regionales y a la propuesta de un enfoque replicable para apoyar el seguimiento educativo a nivel municipal.*
Palabras clave: *Educación Básica; IDEB; Modelado Predictivo; Aprendizaje Automático; Aprendizaje Profundo; Conjuntos Basados en Árboles; Políticas Educativas Basadas en Evidencia.*

1 Introdução

A qualidade da educação básica constitui um dos principais determinantes do desenvolvimento social e econômico, particularmente em países marcados por desigualdades regionais persistentes, como o Brasil (UNESCO, 2022). No contexto brasileiro, tais desigualdades manifestam-se de forma concreta nos indicadores educacionais, sobretudo em unidades federativas com Índice de Desenvolvimento Humano Municipal (IDHM) inferior à média nacional. O estado do Piauí, com IDHM de 0,690, situa-se abaixo da média brasileira (0,766), conforme o Atlas do Desenvolvimento Humano no Brasil (PNUD; IPEA; FJP, 2022), evidenciando assimetrias estruturais que podem repercutir diretamente sobre as condições de oferta educacional e os resultados de aprendizagem.

Nesse cenário, o Índice de Desenvolvimento da Educação Básica (IDEB), criado em 2007, consolidou-se como o principal indicador sintético da qualidade da educação no país. Ao combinar rendimento escolar e desempenho em avaliações padronizadas do Sistema de Avaliação da Educação Básica (SAEB), o IDEB permite monitorar a evolução da educação básica e estabelecer metas nacionais (INEP, 2023). Contudo, por se tratar de um indicador agregado, o IDEB não explicita, de forma direta, os múltiplos fatores estruturais, pedagógicos e socioeconômicos que condicionam suas variações, nem as interações complexas entre essas dimensões.

A literatura educacional tem reiteradamente demonstrado que o desempenho escolar é condicionado por múltiplas dimensões, incluindo qualificação docente, infraestrutura escolar, condições socioeconômicas e investimento público (Hanushek; Woessmann, 2020). Tais dimensões interagem de maneira frequentemente não linear e interdependente, o que limita a capacidade explicativa de modelos baseados exclusivamente em pressupostos de linearidade e independência entre variáveis. Nesse contexto, técnicas de *Machine Learning* (ML) e *Deep Learning* (DL) têm sido progressivamente incorporadas à análise educacional por sua capacidade formal de aproximar funções complexas, modelar interações de alta ordem e capturar padrões não lineares em bases de dados multivariadas e heterogêneas (Hastie; Tibshirani; Friedman, 2009; Chen; Guestrin, 2016).

Apesar desses avanços metodológicos, observa-se que a maior parte das investigações concentra-se em nível individual ou escolar, ou ainda em análises agregadas de abrangência nacional ou estadual. Permanecem incipientes estudos que modelam sistematicamente o IDEB em escala municipal, especialmente nos anos iniciais do ensino fundamental e em contextos regionais caracterizados por maior vulnerabilidade socioeconômica. Ademais, são limitadas as pesquisas que realizam comparação abrangente entre múltiplos algoritmos sob uma estratégia padronizada de avaliação, com controle explícito da variabilidade preditiva.

Diante desse panorama, este estudo tem como objetivos: (i) identificar os fatores associados ao IDEB municipal dos anos iniciais do ensino fundamental no estado do Piauí, a partir de variáveis educacionais, socioeconômicas e financeiras; e (ii) comparar o desempenho de algoritmos de ML e DL na estimativa do IDEB municipal, sob uma estratégia unificada de avaliação que contempla divisão treino/teste e validação cruzada *k-fold*. A contribuição do estudo não consiste em generalizar diretamente os resultados empíricos do Piauí para outros territórios, mas em demonstrar uma abordagem metodológica replicável para modelagem preditiva do IDEB em escala municipal, baseada em dados públicos, seleção sistemática de variáveis, comparação de modelos e interpretação dos preditores. Assim, o recorte piauiense é tratado como estudo de caso aplicado, com potencial de adaptação futura para outras unidades federativas e, posteriormente, para todos os municípios brasileiros.

2 Fundamentação Teórica

A educação básica pública brasileira, assegurada como direito social fundamental pela Constituição Federal de 1988 e regulamentada pela Lei de Diretrizes e Bases da Educação Nacional (Lei n.º 9.394/1996), constitui um dos principais instrumentos para a promoção da cidadania e a redução das desigualdades sociais e regionais. Esses marcos normativos definem a equidade, a qualidade e a valorização dos profissionais da educação como princípios estruturantes do sistema, atribuindo ao Estado a responsabilidade pela garantia do acesso universal, da permanência e do sucesso escolar (Brasil, 1988; Brasil, 1996).

Embora a expansão do acesso à escola pública tenha sido significativa nas últimas décadas, as desigualdades persistem e se manifestam de forma mais acentuada em estados com baixo IDHM, especialmente na região Nordeste. Estudos clássicos e contemporâneos evidenciam que a qualidade educacional resulta de um conjunto de fatores interdependentes, que vão além do rendimento e da aprendizagem aferidos por avaliações padronizadas como o IDEB (Soares; Araújo, 2006). Fatores intraescolares — como a formação inicial e continuada do corpo docente, a estabilidade da equipe, a gestão escolar e a disponibilidade de recursos pedagógicos e infraestrutura — exercem influência direta sobre o desempenho dos estudantes (Silveira; Schneider; Alves, 2023). Por outro lado, as condições socioeconômicas das famílias, como renda, escolaridade parental e capital cultural, impõem limitações às oportunidades educacionais, contribuindo para a reprodução de desigualdades históricas e estruturais.

Esse diagnóstico justifica a necessidade de políticas redistributivas robustas, como o Fundo de Manutenção e Desenvolvimento da Educação Básica e de Valorização dos Profissionais da Educação (FUNDEB) e os parâmetros do Custo Aluno Qualidade Inicial (CAQi) e do Custo Aluno Qualidade (CAQ), que estabelecem padrões mínimos de investimento por aluno e critérios de redistribuição baseados na capacidade fiscal dos entes federados (Carreira; Pinto, 2007; Silveira; Schneider; Alves, 2023). Apesar de avanços atribuíveis a esses mecanismos, persistem disparidades substanciais entre as regiões Sul-Sudeste e Norte-Nordeste, evidenciadas por dados do SAEB e do IDEB (INEP, 2023). Tal persistência reforça a necessidade de análises em escala territorial mais granular, capazes de capturar especificidades locais e orientar políticas públicas mais eficazes.

Nesse cenário, a maior disponibilidade de microdados educacionais e os avanços nas tecnologias de informação têm possibilitado abordagens metodológicas mais sofisticadas para investigar os determinantes do desempenho educacional. Técnicas de ML e DL destacam-se por sua habilidade de lidar com bases de dados multivariadas, além de capturar interações complexas e não lineares entre variáveis, que frequentemente escapam às abordagens estatísticas tradicionais (Zawacki-Richter et al., 2019; Holmes et al., 2019; Chen; Ding, 2023). Essas técnicas vêm sendo aplicadas para prever resultados educacionais, identificar fatores críticos de desempenho, detectar precocemente o risco de evasão escolar e personalizar trajetórias de aprendizagem, conforme demonstrado por estudos nacionais e internacionais recentes (Zawacki-Richter et al., 2019; Holmes et al., 2019; Maia; Bueno; Sato, 2021; Chen; Ding, 2023; Benevento, 2024; Wang; Luo, 2024; Lundberg; Lee, 2017; Molnar, 2020).

A realização de análises nessa escala territorial implica lidar com bases de dados multivariadas e heterogêneas, nas quais múltiplos determinantes educacionais e socioeconômicos interagem de maneira não linear. Nesse contexto, abordagens estatísticas tradicionais podem apresentar limitações para modelar adequadamente tais interações, especialmente quando o objetivo é estimar padrões complexos em nível municipal (Hastie; Tibshirani; Friedman, 2009).

Para além do desempenho preditivo, essas abordagens passaram a incorporar mecanismos formais de interpretabilidade, permitindo examinar a contribuição marginal das variáveis

explicativas nas previsões geradas. Entre esses métodos, destacam-se os valores *SHapley Additive exPlanations* (SHAP), fundamentados na teoria dos jogos cooperativos, que atribuem a cada atributo uma parcela aditiva de contribuição para a predição do modelo (Lundberg; Lee, 2017). Complementarmente, técnicas de seleção e diagnóstico de atributos, como o algoritmo Boruta, o *Variance Inflation Factor* (VIF) e a regressão Lasso com validação cruzada (LassoCV), permitem identificar variáveis estatisticamente relevantes, reduzir redundâncias e mitigar problemas de multicolinearidade, favorecendo modelos mais estáveis e transparentes (Kursa; Rudnicki, 2010; Molnar, 2020). A integração dessas estratégias favorece a consistência analítica e amplia a transparência dos modelos para a formulação de políticas educacionais baseadas em evidências.

Portanto, este trabalho apoia-se em dois pilares teóricos complementares: (i) a literatura sobre os determinantes da qualidade educacional no Brasil, que destaca a influência das dimensões pedagógicas, estruturais, financeiras e socioeconômicas; e (ii) os avanços metodológicos proporcionados por técnicas contemporâneas de ML e DL, que podem contornar algumas limitações das abordagens lineares tradicionais ao incorporar relações mais complexas e fornecer interpretações transparentes dos resultados. Essa integração torna-se ainda mais relevante quando aplicada a contextos regionais marcados por vulnerabilidades históricas, como o do Piauí, contribuindo para diagnósticos mais precisos e para a formulação de estratégias educacionais territorializadas e efetivas.

Dessa forma, após apresentar os fundamentos conceituais que orientam este estudo, a próxima seção discute trabalhos relacionados que empregaram técnicas de ML e DL na análise de indicadores educacionais em diferentes contextos. Em seguida, a Seção 4 detalha a metodologia adotada, abrangendo a organização dos dados, os procedimentos de seleção de variáveis e os modelos preditivos testados. A Seção 5 apresenta os resultados obtidos e a discussão correspondente, articulando desempenho preditivo e interpretabilidade dos modelos. Por fim, a Seção 6 sintetiza as conclusões do estudo e indica direções para pesquisas futuras.

3 Trabalhos Relacionados

Nos últimos anos, a aplicação de técnicas de ML e DL à previsão de indicadores educacionais tem se intensificado, impulsionada pela ampliação do acesso a microdados públicos e pela demanda por diagnósticos quantitativos capazes de subsidiar políticas baseadas em evidências. No contexto brasileiro, observa-se também a consolidação de estudos em Informática na Educação, com aplicações de mineração de dados educacionais e técnicas de ML voltadas à previsão de desempenho acadêmico, conforme evidenciado em trabalhos recentes publicados na Revista Brasileira de Informática na Educação (Souza; Santos, 2021; Rodrigues et al., 2024).

A literatura concentra-se, predominantemente, na previsão do IDEB em escalas agregadas (Maia; Bueno; Sato, 2021), na predição de desempenho acadêmico individual (Chen; Ding, 2023; Wang; Luo, 2024) e na identificação de risco de evasão escolar (Benevento, 2024), empregando diferentes algoritmos supervisionados. Contudo, observa-se predominância de estudos conduzidos em nível escolar ou individual, bem como em escalas nacional ou estadual, sendo ainda limitada a modelagem preditiva sistemática do IDEB em nível municipal, especialmente em contextos de maior vulnerabilidade socioeconômica. Os principais estudos identificados na literatura encontram-se sintetizados na Tabela 1.

No que se refere especificamente à previsão do IDEB, Maia, Bueno e Sato (2021) propuseram modelos preditivos em níveis nacional, regional e municipal, utilizando 61 variáveis educacionais, estruturais e socioeconômicas. Foram avaliados modelos lineares e *Gradient Boosting Machine* (GBM), sendo este último o que apresentou melhor desempenho, com $R^2 \approx$

$0{,}42$, identificando a formação docente como a variável mais relevante. Embora o estudo represente avanço relevante na modelagem preditiva do IDEB, a comparação restringiu-se a um conjunto limitado de algoritmos e não incorporou validação cruzada para estimativa de variabilidade do desempenho, o que limita a análise de estabilidade preditiva dos modelos, conforme sintetizado na Tabela 1.

No âmbito estadual, Soares e Santos (2024) analisaram o IDEB do ensino médio em 222 escolas públicas do Espírito Santo, por meio de regressão linear múltipla. O modelo alcançou um coeficiente de determinação (R^2) ajustado de 0,508 com 13 variáveis explicativas. A análise destacou a taxa de distorção idade-série como fator negativo e o nível socioeconômico como fator positivo. Entretanto, o estudo não realizou comparação entre diferentes técnicas de ML e DL, nem empregou métodos de interpretabilidade para examinar a contribuição marginal dos preditores, além de restringir-se ao nível escolar.

De maneira semelhante, em escala municipal, Rocha, Teixeira e Melo (2015) utilizaram regressão linear múltipla para investigar os fatores associados ao desempenho escolar no Rio Grande do Norte, obtendo $R^2 = 0{,}468$, utilizando microdados do SAEB. O estudo evidenciou maior influência de fatores familiares sobre as notas do SAEB, mas limitou-se a modelos lineares, sem explorar abordagens baseadas em árvores, *ensembles* ou redes neurais, tampouco métodos de explicabilidade.

Além disso, considerando um nível institucional, Benevento (2024) aplicou GBM, *Random Forest* (RF) e regressão logística penalizada (*Logistic Ridge*) para prever risco de evasão escolar em uma escola técnica de Joinville (SC). Entre os modelos testados, o RF apresentou melhor desempenho, com acurácia de 95,28%. Embora desenvolvido em contexto específico do ensino técnico e baseado em microdados individuais, o trabalho demonstra aplicabilidade de algoritmos de ML na previsão de eventos críticos no percurso escolar.

No contexto internacional, Chen e Ding (2023) analisaram o impacto de diferentes modelos de ML na previsão do desempenho acadêmico em escolas da Pensilvânia, testando cinco algoritmos, *Decision Tree* (DT), RF, regressão logística, *Support Vector Machines* (SVM) e rede neural, sendo a rede neural a mais precisa, com acurácia de 60%. De forma complementar, Wang e Luo (2024) propuseram um modelo preditivo para 87 universitários em Wuhan, combinando XGBoost com valores SHAP, alcançando R^2 de 0,82 e Erro Médio Absoluto (MAE) de 0,6. Apesar das diferenças de contexto e unidade de análise em relação ao presente estudo, ambos os trabalhos corroboram a aplicabilidade de modelos baseados em árvores e de abordagens de interpretabilidade em pesquisas educacionais, ainda que em níveis de agregação distintos.

Tabela 1: Síntese comparativa dos trabalhos relacionados sobre predição de indicadores educacionais com ML/DL.

Autores	Alvo	Unidade	Melhor Modelo	Acurácia/R^2
Maia, Bueno e Sato (2021)	IDEB	Município, região e país	GBM	$R^2 = 0{,}42$
Soares e Santos (2024)	IDEB (EM)	Escola	Regressão Linear Múltipla	$R^2 = 0{,}508$
Rocha, Teixeira e Melo (2015)	SAEB	Escola	Regressão Linear Múltipla	$R^2 = 0{,}468$
Benevento (2024)	Evasão	Escola	RF	95,28%
Chen e Ding (2023)	Desempenho	Aluno	ANN	60%
Wang e Luo (2024)	Desempenho	Aluno	XGBoost	$R^2 = 0{,}82$

Dessa forma, a partir da análise comparativa apresentada na Tabela 1, a literatura evidencia três lacunas centrais: (i) a escassez de estudos voltados aos anos iniciais do ensino fundamental em escala municipal, especialmente em contextos de maior vulnerabilidade socioeconômica; (ii) a ausência de comparações sistemáticas entre múltiplos algoritmos de ML e DL utilizando

validação cruzada como estimativa explícita de variabilidade preditiva; e (iii) a limitada incorporação de métodos de interpretabilidade capazes de analisar a contribuição marginal dos preditores em indicadores educacionais agregados.

Nesse cenário, o presente estudo diferencia-se ao concentrar-se nos municípios do estado do Piauí, região marcada por heterogeneidade socioeconômica, e ao comparar 17 modelos de ML e DL sob estratégia unificada de avaliação, com divisão treino/teste e validação cruzada *k-fold*. Ademais, integra um *pipeline* sistemático de seleção de variáveis à análise interpretativa por SHAP e importância por permutação. Sua contribuição reside na articulação entre comparação metodológica abrangente, controle explícito da variabilidade estatística e interpretação estruturada dos resultados, não como evidência de superioridade territorial do caso piauiense, mas como demonstração de uma estratégia analítica que pode ser replicada e testada em outros contextos federativos.

4 Metodologia

4.1 Organização e Tratamento dos Dados

Os dados utilizados neste estudo foram obtidos exclusivamente de fontes públicas oficiais, garantindo rastreabilidade e transparência metodológica. Foram integrados microdados do Censo Escolar, do SAEB e do IDEB, disponibilizados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), bem como informações socioeconômicas e demográficas do Instituto Brasileiro de Geografia e Estatística (IBGE), dados fiscais do Sistema de Informações sobre Orçamentos Públicos em Educação (SIOPE), do Sistema de Informações Contábeis e Fiscais do Setor Público Brasileiro (SICONFI) e registros do SAGICAD.

No que se refere à delimitação analítica, a unidade de análise adotada foi o município, contemplando os 224 municípios do estado do Piauí, nos anos ímpares compreendidos entre 2013 e 2023, em consonância com o calendário oficial de divulgação do IDEB. A base consolidada inicialmente continha 1.344 observações e 128 variáveis.

A partir dessa consolidação inicial, o processo de organização dos dados envolveu etapas de limpeza, agregação e tratamento das variáveis. Inicialmente, foram removidas variáveis com elevada proporção de valores ausentes, como “Nota SAEB em Matemática” e “Nota SAEB em Língua Portuguesa”. Além da alta incompletude, essas variáveis já possuíam suas informações incorporadas na variável agregada “Nota SAEB – Nota Média Padronizada (N)”, construída a partir dos mesmos componentes, o que tornava sua permanência redundante. Adicionalmente, foram excluídas as observações que apresentavam ausência da variável dependente (IDEB), uma vez que tais registros inviabilizam a estimação dos modelos preditivos. Após essa etapa, a base passou a conter 1.262 observações e 126 variáveis.

No que diz respeito à estrutura dos microdados, os microdados do Censo Escolar, originalmente estruturados em nível de escola, foram agregados ao nível municipal. Variáveis quantitativas, como número de matrículas, docentes, salas e equipamentos, foram somadas por município e ano. Para variáveis binárias indicativas da presença de recursos escolares (água potável, energia elétrica, biblioteca, laboratório, quadra esportiva, acessibilidade, conselho escolar, entre outras), adotou-se a soma das escolas que apresentavam o respectivo recurso, preservando a variação intra-municipal e evitando perda de informação decorrente da utilização exclusiva de proporções médias.

Adicionalmente, foi construída a variável “quantidade de escolas”, correspondente ao total de estabelecimentos públicos por município e ano, utilizada como controle estrutural do porte da

rede municipal nas análises subsequentes, permitindo distinguir efeitos associados à escala da rede daqueles relacionados à disponibilidade relativa de recursos. No tocante às variáveis sem atualização disponível para 2023, foram adotadas estratégias de projeção de curto prazo. A população municipal foi estimada por crescimento geométrico, calculando-se a taxa de variação entre 2021 e 2022 e aplicando-a ao valor observado em 2022, conforme:

$$P_{2023} = P_{2022}(1 + r) \quad (1)$$

Em que:

$$r = \frac{P_{2022} - P_{2021}}{P_{2021}} \quad (2)$$

Para o Produto Interno Bruto (PIB) municipal e seus componentes, ajustou-se uma regressão linear simples aos dados disponíveis no período de 2010 a 2021, utilizando o ano como variável explicativa. A partir do modelo estimado, foram projetados os valores de 2022 e 2023. O procedimento foi aplicado ao PIB total, PIB per capita, Valor Adicionado Bruto (VAB) setorial e impostos líquidos.

Após a consolidação e limpeza inicial da base, procedeu-se à imputação dos valores faltantes remanescentes nas variáveis explicativas, que apresentavam pequenas quantidades de ausência (entre 4 e 15 registros, a depender da variável). A imputação foi realizada pela mediana, escolhida por sua menor sensibilidade a *outliers* e assimetrias. Em seguida, realizou-se a integração das variáveis provenientes das diferentes fontes, utilizando os códigos dos municípios do IBGE e o ano como chaves de pareamento. Por fim, os valores foram padronizados (*z-score*) nos modelos sensíveis à escala, enquanto essa etapa foi dispensada para algoritmos invariantes, como regressão linear, DT, RF, GBM e XGBoost.

4.2 Seleção de Variáveis

Com a finalidade de obter um conjunto de atributos representativo, parcimonioso e livre de redundâncias, adotou-se um processo sistemático de seleção de variáveis. A base de dados utilizada neste estudo era originalmente composta por 128 atributos, abrangendo informações educacionais, socioeconômicas e financeiras referentes aos municípios do estado do Piauí. Desse conjunto inicial, seis variáveis foram removidas por não constituírem preditores adequados para a modelagem: a variável resposta (IDEB), as notas do SAEB em Matemática e Português — já representadas pela variável Nota SAEB padronizada (N) — e os identificadores administrativos (ano, nome e código do município). Após essa etapa, permaneceram 122 variáveis explicativas.

Na sequência, foram excluídas variáveis que compõem diretamente a fórmula do cálculo do IDEB, a fim de evitar vazamento de informação e redundância estrutural. Essa remoção reduziu o conjunto para 119 variáveis explicativas, conforme ilustrado na Figura 1. Tal procedimento foi necessário para assegurar que os modelos não incorporassem componentes diretamente vinculados ao indicador-alvo, o que poderia inflar artificialmente o desempenho preditivo e comprometer a validade inferencial dos resultados.

A partir desse conjunto, aplicou-se um *pipeline* integrado de seleção de atributos, composto por quatro métodos complementares, também sintetizado na Figura 1, envolvendo quatro técnicas complementares de seleção de atributos. Inicialmente, realizou-se análise de

multicolinearidade por meio do VIF, adotando-se como critério de permanência valores inferiores ou iguais a 10. A aplicação iterativa desse procedimento resultou na redução do conjunto para 38 variáveis. Em seguida, empregou-se a regressão LassoCV, cuja penalização L_1, induz esparsidade ao reduzir coeficientes pouco informativos a zero, reduzindo o conjunto para 15 variáveis.

Posteriormente, procedeu-se à análise de importância baseada em valores SHAP, os quais mensuram a contribuição marginal média de cada atributo para a predição. Considerando-se um limiar de importância média igual ou superior a 0,01, com isso reduziu o conjunto para 12 variáveis. Por fim, aplicou-se o algoritmo Boruta, baseado em RF, que compara a importância das variáveis reais com atributos aleatorizados ("*shadow features*"), retendo apenas aquelas cujo poder explicativo supera o ruído estatístico. O procedimento resultou em seis variáveis. A convergência entre os métodos aplicados consolidou o conjunto final de seis variáveis explicativas, utilizado como entrada para todos os modelos de ML e DL avaliados neste estudo, conforme a Figura 1.

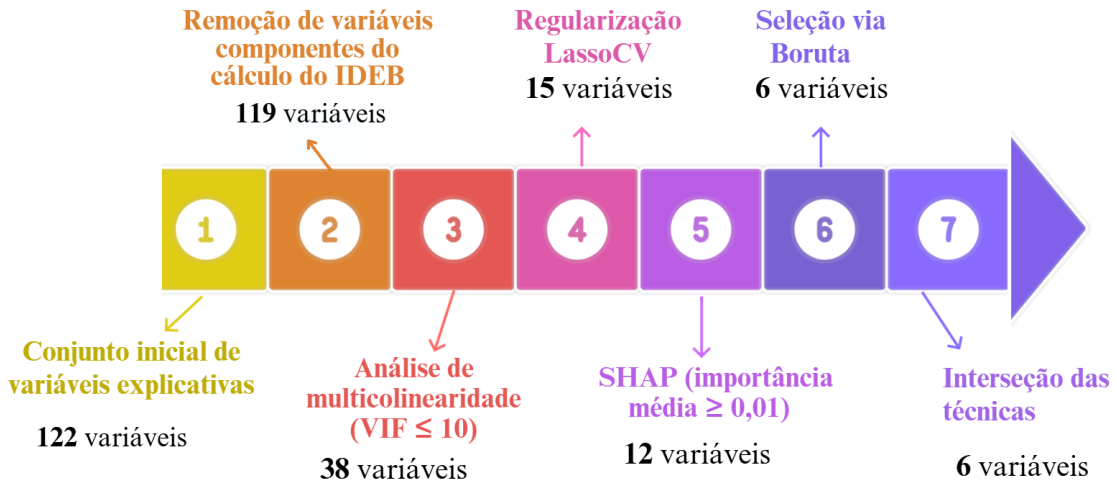

Figura 1: *Pipeline* sequencial de seleção das variáveis explicativas.

As seis variáveis selecionadas foram: PIB per capita, taxa de distorção idade-série, percentual de docentes sem formação superior, área plantada das lavouras temporárias e permanentes, valor da produção vegetal e repasses do programa Criança Feliz. Esse conjunto de variáveis sintetiza dimensões educacionais e socioeconômicas que apresentaram maior poder explicativo na variação do IDEB entre os municípios do estado do Piauí.

Importa esclarecer que o processo descrito não consistiu apenas em uma filtragem sequencial automática de atributos até a obtenção do conjunto final. Em cada etapa do *pipeline* — correspondentes aos subconjuntos com 119, 38, 15, 12 e 6 variáveis — os 17 modelos preditivos avaliados neste estudo (10 de ML e 7 de DL) foram integralmente reestimados. Para cada subconjunto de variáveis, procedeu-se à nova divisão treino-teste e à validação cruzada, calculando-se novamente as métricas de desempenho (R^2, MAE e RMSE). Dessa forma, cada estágio de redução dimensional foi submetido a avaliação empírica independente, permitindo comparar diretamente o impacto do número de variáveis sobre a capacidade preditiva dos modelos.

Essa estratégia possibilitou comparar o desempenho dos modelos em cada estágio de redução dimensional. Verificou-se que a diminuição do número de variáveis não comprometeu a capacidade preditiva e que o subconjunto final com seis variáveis apresentou o melhor desempenho médio entre os conjuntos avaliados, considerando as métricas de validação adotadas.

Assim, a definição do conjunto final resultou da combinação entre critérios estatísticos de seleção e evidência empírica de desempenho observada ao longo do *pipeline*.

4.3 Modelagem Preditiva

A estratégia de modelagem preditiva foi estruturada com o propósito de comparar diferentes abordagens supervisionadas aplicadas à predição do IDEB municipal, contemplando modelos lineares, métodos baseados em instâncias, algoritmos baseados em árvores e arquiteturas neurais profundas. A seleção dos algoritmos considerou três critérios principais: (i) adequação a dados tabulares heterogêneos; (ii) capacidade de capturar relações lineares e não lineares entre as variáveis explicativas e o IDEB; e (iii) comparabilidade entre modelos com distintos níveis de complexidade estrutural.

No grupo de ML, foram implementados Regressão Linear Múltipla e regressões penalizadas, *K-Nearest Neighbors* (KNN), DT e modelos de *ensemble* como RF, GBM e XGBoost. Os modelos lineares e penalizados permitem modelar relações lineares e mitigar efeitos de multicolinearidade; o KNN opera com base na similaridade entre instâncias; e os métodos baseados em árvores e *ensembles* capturam interações e relações não lineares entre atributos, sendo particularmente adequados a dados tabulares.

No âmbito do DL, foram avaliadas Redes Neurais Artificiais (ANN) do tipo MLP, Redes Neurais Convolucionais (CNN), Redes Neurais Recorrentes (RNN) com unidades de memória do tipo *Long Short-Term Memory* (LSTM) e uma arquitetura híbrida CNN+LSTM. As arquiteturas MLP utilizadas possuem múltiplas camadas densas treinadas por retropropagação, caracterizando-se como redes profundas no sentido arquitetural do termo. As CNN foram incluídas para avaliar a extração hierárquica de padrões a partir das entradas estruturadas, enquanto as LSTM permitiram testar empiricamente se mecanismos internos de memória contribuiriam para representar possíveis variações associadas aos anos analisados, ainda que os dados não configurem uma série temporal longitudinal clássica. A arquitetura híbrida combinou operações convolucionais e recorrentes, permitindo avaliar o comportamento integrado dessas estruturas frente ao mesmo conjunto de atributos.

Dessa forma, as arquiteturas de DL foram incluídas não como substitutas presumidamente superiores aos modelos tradicionais, mas como alternativas empíricas para avaliar se estruturas densas, convolucionais, recorrentes e híbridas seriam capazes de capturar padrões não lineares adicionais nos dados tabulares agregados. Essa escolha também permitiu examinar criticamente os limites dessas arquiteturas em uma base municipal de tamanho moderado, composta por variáveis agregadas e sem estrutura temporal densa.

Os dados foram padronizados por meio de normalização *z-score* nos modelos sensíveis à escala das variáveis (regressões penalizadas, KNN, MLP e Deep MLP), de modo a evitar que diferenças de magnitude influenciassem coeficientes ou gradientes de otimização. Para os algoritmos invariantes à escala, Regressão Linear Múltipla, DT, RF, GBM e XGBoost, a normalização não foi aplicada, dado que essas técnicas operam por partições ou combinações de variáveis em termos absolutos.

Com o objetivo de assegurar robustez estatística, capacidade de generalização e comparabilidade entre os modelos, adotou-se uma estratégia padronizada de avaliação entre os modelos de ML e DL. A base de dados foi particionada em 70% para treino e 30% para teste. No conjunto de treino, aplicou-se validação cruzada do tipo *k-fold (k = 5)* para todos os modelos, permitindo estimar a variabilidade do desempenho preditivo e reduzir o risco de sobreajuste. O método estatístico adotado foi idêntico para ML e DL, consistindo na divisão do conjunto de

treino em cinco partições mutuamente exclusivas, nas quais, a cada iteração, quatro *folds* foram utilizados para ajuste do modelo e um *fold* para validação.

Nos modelos de ML implementados no *scikit-learn*, a validação cruzada foi realizada por meio das funções nativas da biblioteca (*cross_val_score*), que automatizam o particionamento e a reestimação do modelo em cada *fold*. Já nos modelos de DL, implementados em *Keras/TensorFlow*, o mesmo procedimento estatístico foi operacionalizado manualmente utilizando a classe *KFold* do *scikit-learn* para gerar os índices de divisão dos *folds*. Em cada iteração, a arquitetura da rede foi reconstruída e treinada novamente com inicialização aleatória dos pesos, garantindo independência entre as execuções e evitando reutilização de parâmetros previamente ajustados. Assim, o método estatístico de validação cruzada foi idêntico para ML e DL, diferindo apenas o mecanismo computacional de implementação.

Em ambos os paradigmas, as métricas de desempenho consideradas foram MAE, Raiz do Erro Quadrático Médio (RMSE) e R^2, calculadas para validação cruzada, treino final e teste. Essa padronização permitiu avaliar não apenas o desempenho médio, mas também a variabilidade associada às estimativas, fornecendo base empírica consistente para comparação entre os modelos.

Adicionalmente, realizou-se uma comparação estatística transversal entre os 17 modelos avaliados. Para isso, os valores obtidos nos cinco *folds* da validação cruzada foram tratados como medidas pareadas de desempenho para cada métrica. Inicialmente, aplicou-se o teste não paramétrico de Friedman para verificar a existência de diferenças globais entre os modelos quanto ao R^2, MAE e RMSE. Em seguida, quando identificada diferença global significativa, foram conduzidas comparações *post-hoc* par-a-par por meio do teste de Wilcoxon pareado, com correção de Holm para múltiplas comparações. Esse procedimento permitiu avaliar a existência de diferenças gerais entre os algoritmos e examinar, de forma controlada, possíveis diferenças específicas entre pares de modelos.

4.4 Ferramentas Computacionais e Ajuste de Hiperparâmetros

As análises foram implementadas em linguagem *Python*, utilizando a biblioteca *Scikit-learn* para os modelos de ML e *Keras*, com *TensorFlow* como *backend*, para as arquiteturas de DL. O pré-processamento dos dados foi realizado com as bibliotecas *Pandas* e *NumPy*, e as visualizações foram produzidas com *Matplotlib* e *Seaborn*. A Tabela 2 apresenta os algoritmos avaliados e os respectivos hiperparâmetros selecionados. No grupo de ML, foram considerados modelos lineares e penalizados, métodos baseados em instâncias, árvores de decisão e técnicas de *ensemble*. No grupo de DL, foram implementadas sete arquiteturas neurais, incluindo MLP, variações com regularização (*Dropout* e *Batch Normalization*), além de CNN, LSTM e híbridas.

A definição dos hiperparâmetros dos modelos de ML foi realizada por meio de busca em grade (*Grid Search*) com validação cruzada *k-fold (k = 5)* exclusivamente no conjunto de treinamento, adotando como critério de seleção o maior R^2 na validação cruzada. Para cada algoritmo, foram exploradas combinações sistemáticas de parâmetros estruturais. No caso do XGBoost, por exemplo, foram avaliadas combinações de $n\_estimators \in \{200,500,1000\}$, $max\_depth \in \{2,3,4\}$, $learning\_rate \in \{0.01, 0.05\}$, $reg\_alpha \in \{0, 1\}$ e $reg\_lambda \in \{1, 5, 10\}$. Para RF, variaram-se *n_estimators*, *max_depth* e *min_samples_split*; para SVR, foram testadas combinações de $C \in \{0.1,1,10\}$ e $epsilon \in \{0.01, 0.1\}$. Após a identificação da melhor configuração, cada modelo foi reestimado integralmente no conjunto de treinamento antes da avaliação no conjunto de teste.

Para as arquiteturas de DL, os modelos foram treinados por até 200 épocas, com *batch size* igual a 32 e aplicação de *early stopping* ($patience = 20$), monitorando a perda de validação.

Diferentes configurações estruturais (número de camadas, neurônios por camada, taxas de *dropout* e *learning rate*) foram comparadas sob o mesmo protocolo de validação cruzada aplicado aos modelos de ML. Embora não tenha sido conduzida busca exaustiva em grade para todas as combinações arquiteturais, em razão do custo computacional, as configurações avaliadas foram definidas com base em evidências da literatura e comparadas sistematicamente quanto ao desempenho médio e à estabilidade entre *folds*.

Os *scripts* utilizados para pré-processamento, seleção de variáveis, modelagem preditiva e teste estatístico, bem como a base de dados tratada, encontram-se disponíveis em repositório público, com o objetivo de garantir transparência, reprodutibilidade e auditoria metodológica dos experimentos. O material pode ser acessado em: https://github.com/mariaeva020/ideb-piaui-ml-dl-/.

Tabela 2: Principais hiperparâmetros dos algoritmos de ML e DL aplicados neste estudo.

Grupo	Modelo / Arquitetura	Principais Hiperparâmetros
ML	Regressão Linear Múltipla	—
	Lasso	$\alpha = 0.1$
	Ridge	$\alpha = 1.0$
	Elastic Net	$\alpha = 0.1;\ l1\_ratio = 0.5$
	KNN	$n\_neighbors = 7$
	SVR	$kernel = 'rbf';\ C = 1.0;\ epsilon = 0.1$
	DT	$max\_depth = 3;\ min\_samples\_split = 15;$ $min\_samples\_leaf = 10$
	RF	$n\_estimators = 300;\ max\_depth = 4;\ min\_samples\_split =$ $10;\ min\_samples\_leaf = 10;\ max\_features = 'sqrt'$
	GBM	$n\_estimators = 200;\ learning\_rate = 0.01;\ max\_depth =$ $3;\ subsample = 0.8;\ min\_samples\_leaf = 5$
	XGBoost	$n\_estimators = 1000;\ learning\_rate = 0.01;\ max\_depth =$ $2;\ subsample = 0.8;\ colsample\_bytree = 0.8;\ reg\_alpha =$ $1.0;\ reg\_lambda = 10.0$
DL	MLP	64–32 neurônios; ativação ReLU nas camadas ocultas; saída linear; otimizador $Adam;\ learning\_rate = 0.001$
	MLP com Dropout	128–64 neurônios; $dropout = 0.3$ após cada camada oculta; ativação $ReLU$; saída linear; $Adam;\ learning\_rate = 0.001$
	MLP com Batch Normalization	128–64 neurônios; $BatchNorm$ após cada camada oculta; ativação $ReLU$; saída linear; $Adam;\ learning\_rate = 0.001$
	Deep MLP	256–128–64–32 neurônios; $dropout = 0.4\ e\ 0.3$ em camadas intermediárias; ativação $ReLU$; saída linear; $Adam;$ $learning\_rate = 0.0005$
	CNN	$Conv1D: filters = 32;\ kernel\_size = 2;\ MaxPooling1D;$ camada densa final com 50 neurônios e ativação $ReLU$; saída linear; $Adam;\ learning\_rate = 0.001$
	LSTM	50 unidades LSTM; ativação $tanh$; saída linear; $Adam;$ $learning\_rate = 0.001$
	CNN+LSTM	$Conv1D$: 64 filtros; $kernel\_size = 2;\ MaxPooling1D$; 50 unidades LSTM; ativação $tanh$; saída linear; $Adam;$ $learning\_rate = 0.001$

5 Resultados e Discussão

5.1 Desempenho dos Modelos de ML

Os resultados indicam que o XGBoost apresentou o melhor desempenho médio entre os modelos de ML avaliados. No conjunto de teste, o modelo alcançou $R^2 = 0{,}5542$, MAE de $0{,}5050$ e RMSE de 0,6431. A validação cruzada corroborou esse desempenho, com $R^2 = 0{,}5455 \pm$

$0{,}0293$, evidenciando consistência entre as partições amostrais. Embora esse valor indique o melhor desempenho entre os modelos avaliados, o R^2 deve ser interpretado como capacidade explicativa moderada, compatível com a natureza multifatorial do desempenho educacional municipal. A variabilidade não explicada pode estar relacionada à ausência de dimensões pedagógicas, institucionais, familiares e territoriais não mensuradas nas bases públicas utilizadas, bem como à limitação temporal da série analisada.

Outros algoritmos baseados em árvores também apresentaram desempenhos satisfatórios, embora inferiores ao XGBoost. O RF obteve $R^2 = 0{,}5222$ no teste e $0{,}5311 \pm 0{,}0292$ na validação cruzada, enquanto o GBM registrou $R^2 = 0{,}5123$ no teste e $0{,}5270 \pm 0{,}0256$ na validação cruzada. Embora não tenham superado o XGBoost, esses modelos demonstraram capacidade de capturar relações não lineares e interações entre variáveis, mantendo estabilidade preditiva.

Os modelos lineares exibiram desempenho inferior, ainda que consistente. A regressão linear múltipla apresentou $R^2 = 0{,}5037$ no teste e $0{,}5227 \pm 0{,}0351$ na validação cruzada. As versões penalizadas produziram resultados próximos, com menor variação entre treino e validação, mas limitada capacidade de modelar interações complexas. Observa-se que, embora o ajuste no treino seja satisfatório, os modelos lineares tendem a apresentar menor desempenho no teste quando comparados aos métodos baseados em árvores. Os resultados completos dos modelos avaliados encontram-se sintetizados na Tabela 3.

Tabela 3: Desempenho dos modelos de ML para os anos iniciais do ensino fundamental com as variáveis selecionadas.

Modelo	Métricas	Treino	Teste	Validação Cruzada
Regressão Linear	R^2	0.5336	0.5037	0.5227 ± 0.0351
	MAE	0.4968	0.5240	0.5014 ± 0.0180
	RMSE	0.6368	0.6785	0.6430 ± 0.0261
Lasso	R^2	0.5063	0.4575	0.5027 ± 0.0308
	MAE	0.5084	0.5447	0.5099 ± 0.0177
	RMSE	0.6552	0.7094	0.6565 ± 0.0259
Ridge	R^2	0.5336	0.5037	0.5228 ± 0.0351
	MAE	0.4968	0.5240	0.5014 ± 0.0180
	RMSE	0.6368	0.6785	0.6429 ± 0.0261
Elastic Net	R^2	0.5214	0.4835	0.5171 ± 0.0310
	MAE	0.5007	0.5321	0.5023 ± 0.0192
	RMSE	0.6451	0.6922	0.6469 ± 0.0237
RF	R^2	0.6012	0.5222	0.5311 ± 0.0292
	MAE	0.4548	0.5192	0.4917 ± 0.0245
	RMSE	0.5889	0.6657	0.6373 ± 0.0193
GBM	R^2	0.6102	0.5123	0.5270 ± 0.0256
	MAE	0.4535	0.5254	0.4973 ± 0.0211
	RMSE	0.5821	0.6726	0.6403 ± 0.0219
XGBoost	**R^2**	**0.6385**	**0.5542**	**0.5455 ± 0.0293**
	MAE	**0.4354**	**0.5050**	**0.4867 ± 0.0197**
	RMSE	**0.5606**	**0.6431**	**0.6275 ± 0.0216**
DT	R^2	0.5400	0.4275	0.4864 ± 0.0361
	MAE	0.4944	0.5628	0.5224 ± 0.0220
	RMSE	0.6324	0.7288	0.6669 ± 0.0216
SVR	R^2	0.6021	0.5170	0.5350 ± 0.0434
	MAE	0.4414	0.5206	0.4891 ± 0.0289
	RMSE	0.5882	0.6694	0.6343 ± 0.0288
KNN	R^2	0.6263	0.4622	0.4749 ± 0.0701
	MAE	0.4495	0.5492	0.5290 ± 0.0324
	RMSE	0.5700	0.7063	0.6730 ± 0.0387

Conforme evidenciado na Tabela 3, os melhores valores foram destacados em negrito, considerando que maiores valores de R² e valores menores de MAE e RMSE indicam melhor desempenho. A comparação entre os algoritmos indica o destaque relativo do XGBoost nos conjuntos de avaliação, incluindo treino, teste e validação cruzada. Esses resultados indicam a adequação do XGBoost, enquanto método de *boosting* baseado em árvores, para a tarefa preditiva analisada, especialmente diante da heterogeneidade das variáveis e da possibilidade de interações não lineares entre os preditores.

5.2 Desempenho dos Modelos de DL

Nas arquiteturas DL avaliadas, a MLP apresentou o melhor desempenho médio. No conjunto de teste, obteve $R^2 = 0{,}5462$, MAE de $0{,}5210$ e RMSE de $0{,}6708$. Na validação cruzada, registrou $R^2 = 0{,}5205 \pm 0{,}0470$, indicando desempenho competitivo, embora com maior variabilidade em comparação ao XGBoost. Entre as demais arquiteturas, a CNN 1D apresentou resultados próximos aos da MLP, com $R^2 = 0{,}4939$ no teste e $0{,}5050 \pm 0{,}0346$ na validação cruzada, sugerindo capacidade razoável de capturar padrões estruturais nas entradas tabulares reorganizadas. A LSTM alcançou $R^2 = 0{,}5291$ no teste, porém apresentou maior oscilação na validação cruzada ($0{,}4557 \pm 0{,}0467$), evidenciando sensibilidade a variações amostrais. A arquitetura híbrida CNN + LSTM também não superou a MLP.

As variações da MLP indicaram que o aumento da complexidade arquitetural não resultou necessariamente em melhora de desempenho. A MLP com *Dropout* apresentou redução no R² de teste ($0{,}4930$) e maior erro médio. A versão com *Batch Normalization* demonstrou instabilidade significativa na validação cruzada ($0{,}1332 \pm 0{,}2001$), sugerindo inadequação da arquitetura ao tamanho e natureza do conjunto de dados. O Deep MLP apresentou indícios de sobreajuste, com desempenho inferior no teste e na validação cruzada. A Tabela 4 consolida os resultados das arquiteturas de DL. De forma geral, embora a MLP tenha apresentado desempenho competitivo, nenhuma arquitetura de DL superou o XGBoost em termos de R^2, MAE ou RMSE.

Tabela 4: Desempenho dos modelos de DL para os anos iniciais do ensino fundamental com as variáveis selecionadas.

Modelo	Métricas	Treino	Teste	Validação Cruzada
MLP	**R²**	**0.6038**	**0.5462**	**0.5205 ± 0.0470**
	MAE	**0.4582**	**0.5210**	**0.4989 ± 0.0193**
	RMSE	**0.5869**	**0.6708**	**0.6413 ± 0.0243**
MLP com Dropout	R²	0.5310	0.4930	0.4570 ± 0.0603
	MAE	0.4896	0.5270	0.5280 ± 0.0124
	RMSE	0.6386	0.6858	0.6819 ± 0.0201
MLP com BatchNorm	R²	0.5335	0.3755	0.1332 ± 0.2001
	MAE	0.4745	0.5784	0.6323 ± 0.0433
	RMSE	0.6369	0.7612	0.8557 ± 0.0878
Deep MLP	R²	0.4732	0.3829	0.3943 ± 0.0935
	MAE	0.5151	0.5603	0.5497 ± 0.0237
	RMSE	0.6768	0.7566	0.7185 ± 0.0198
CNN 1D	R²	0.5386	0.4939	0.5050 ± 0.0346
	MAE	0.4985	0.5474	0.5114 ± 0.0280
	RMSE	0.6334	0.6852	0.6528 ± 0.0342
LSTM	R²	0.5114	0.5291	0.4557 ± 0.0467
	MAE	0.5072	0.5353	0.5405 ± 0.0207
	RMSE	0.6484	0.6833	0.6836 ± 0.0238
CNN + LSTM	R²	0.5441	0.5036	0.5107 ± 0.0368
	MAE	0.4953	0.5441	0.5087 ± 0.0266
	RMSE	0.6296	0.6786	0.6487 ± 0.0308

Esses resultados indicam que, para a base tabular e agregada utilizada neste estudo, os modelos baseados em árvores apresentaram melhor desempenho que as arquiteturas de DL. As redes neurais foram avaliadas para verificar se estruturas densas, convolucionais, recorrentes e híbridas poderiam capturar padrões adicionais nos dados; entretanto, seu desempenho foi limitado pelo tamanho da amostra, pela agregação municipal das variáveis e pela ausência de sequências temporais densas. Assim, os achados não indicam inadequação do DL em problemas educacionais, mas mostram que, neste recorte, o XGBoost apresentou melhor equilíbrio entre desempenho preditivo, estabilidade e interpretabilidade.

5.3 Comparação entre os Melhores Modelos

Após a avaliação individual dos modelos de ML e DL, procedeu-se à comparação direta entre aqueles que apresentaram melhor desempenho médio na validação cruzada: o XGBoost, no conjunto de modelos de ML, e a MLP, entre as arquiteturas de DL. A comparação considerou as métricas R^2, MAE e RMSE, avaliadas tanto na validação cruzada com cinco *folds (k = 5)* quanto no conjunto de teste.

Em termos de R^2, o XGBoost obteve desempenho ligeiramente superior ao da MLP. No conjunto de teste, o XGBoost alcançou $R^2 = 0{,}5542$, enquanto a MLP atingiu $R^2 = 0{,}5462$. Na validação cruzada, os valores médios foram $R^2 = 0{,}5455 \pm 0{,}0293$ para o XGBoost e $R^2 = 0{,}5205 \pm 0{,}0470$ para a MLP. Além da média mais elevada, o XGBoost exibiu menor dispersão entre os *folds*, indicando maior estabilidade preditiva.

Resultados semelhantes foram observados para o MAE. No conjunto de teste, o XGBoost registrou $MAE = 0{,}5050$, enquanto a MLP apresentou $MAE = 0{,}5210$. Na validação cruzada, os valores médios foram $0{,}4867 \pm 0{,}0197$ para o XGBoost e $0{,}4989 \pm 0{,}0193$ para a MLP. Esses resultados indicam que, em média, o XGBoost produziu previsões mais próximas dos valores observados do IDEB. Considerando o RMSE, métrica que penaliza mais fortemente erros de maior magnitude, o XGBoost manteve desempenho superior.

No conjunto de teste, obteve $RMSE = 0{,}6431$, enquanto a MLP alcançou $RMSE = 0{,}6708$. Na validação cruzada, os valores médios foram $0{,}6275 \pm 0{,}0216$ para XGBoost e $0{,}6413 \pm 0{,}0243$ para MLP. Além do menor erro médio, o XGBoost apresentou menor dispersão, sugerindo maior estabilidade preditiva. De forma consolidada, a Figura 2 sintetiza essas diferenças ao apresentar as médias e os desvios-padrão das três métricas na validação cruzada, permitindo visualizar simultaneamente o desempenho médio e a estabilidade dos modelos.

Embora a Figura 2 apresente a comparação visual entre os modelos de melhor desempenho em cada grupo, realizou-se também uma análise estatística transversal contemplando todos os 17 modelos avaliados. Para isso, os resultados obtidos nos cinco *folds* da validação cruzada foram utilizados como medidas pareadas de desempenho, considerando separadamente as métricas R^2, MAE e RMSE.

Inicialmente, aplicou-se o teste não paramétrico de Friedman, cuja hipótese nula assume que todos os modelos apresentam desempenho equivalente, isto é, que não há diferença estatisticamente significativa entre os *rankings* dos algoritmos nos *folds* da validação cruzada. A hipótese alternativa, por sua vez, indica que pelo menos um modelo apresenta desempenho estatisticamente distinto dos demais. O teste de Friedman indicou diferença estatisticamente significativa entre os modelos nas três métricas avaliadas. Para o R^2, obteve-se estatística de Friedman igual a 55,8018, com $p = 0{,}000003$. Para o MAE, a estatística foi 60,3629, com $p < 0{,}000001$. Para o RMSE, a estatística foi 55,8018, com $p = 0{,}000003$. Assim, rejeitou-se a

hipótese nula de equivalência global entre os modelos, indicando que os algoritmos não apresentaram desempenho equivalente quando avaliados conjuntamente na validação cruzada.

Como o teste de Friedman indica apenas a existência de diferença global, mas não identifica quais pares de modelos diferem entre si, foram realizadas comparações *post-hoc* par-a-par. Para isso, utilizou-se o teste de Wilcoxon pareado, adequado ao pareamento dos *folds*, com correção de Holm para múltiplas comparações. Nesse teste, a hipótese nula considera que não há diferença estatisticamente significativa entre o desempenho de cada par de modelos comparado. Após o ajuste dos p-valores, nenhuma comparação específica permaneceu estatisticamente significativa ao nível de 5%, inclusive entre XGBoost e MLP. Esse resultado indica que, embora exista diferença global entre os modelos, os dados disponíveis não fornecem evidência estatística suficiente para afirmar que um par específico de algoritmos difere significativamente após a correção para múltiplas comparações. Esse achado deve ser interpretado com cautela, pois o número reduzido de *folds* e a elevada quantidade de comparações reduzem o poder estatístico dos testes *post-hoc*.

Apesar dessa limitação inferencial, a análise dos rankings médios derivados do teste de Friedman reforçou o melhor desempenho relativo do XGBoost. Nessa análise, os modelos são ordenados em cada *fold* conforme o critério da métrica avaliada: maiores valores de R^2 e menores valores de MAE e RMSE correspondem a melhores posições. Assim, menores *rankings* médios indicam melhor desempenho relativo entre os algoritmos avaliados. O XGBoost apresentou o menor *ranking* médio para R^2 e RMSE, ambos iguais a 2,6, e o menor *ranking* médio para MAE, igual a 2,0. Portanto, a escolha do XGBoost como modelo final decorre de seu melhor desempenho relativo, menor erro médio e maior estabilidade entre os *folds*, e não de uma superioridade estatística absoluta sobre todos os demais modelos.

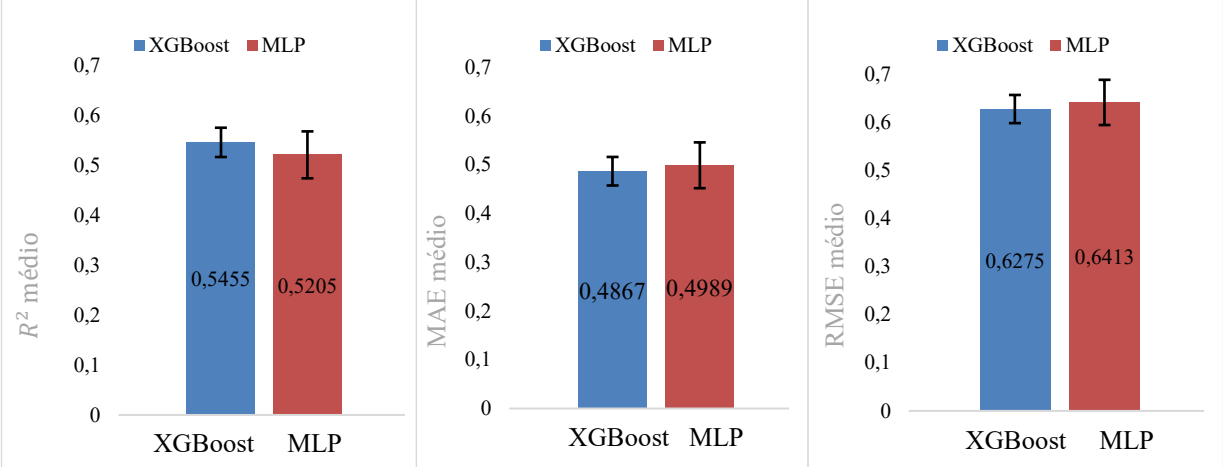

Figura 2: Comparação entre XGBoost e MLP na validação cruzada (k = 5), considerando médias e desvios-padrão para R^2, MAE e RMSE.

Em comparação com o estudo de Maia, Bueno e Sato (2021), que utilizou modelos lineares e GBM para previsão do IDEB em nível nacional, os resultados deste estudo devem ser interpretados com cautela, pois os recortes territoriais, as bases de dados, os algoritmos e os protocolos de validação não são diretamente equivalentes. Ainda assim, observa-se que o XGBoost apresentou desempenho competitivo no contexto piauiense, com $R^2 = 0{,}5542$ no conjunto de teste, além de menores valores de MAE e RMSE em relação aos reportados naquele estudo para os anos iniciais. Diferentemente do estudo citado, esta pesquisa incorporou validação cruzada *k-fold (k = 5)*, o que permitiu estimar a variabilidade do desempenho entre partições amostrais. Dessa forma, a comparação não deve ser entendida como demonstração de

superioridade generalizável, mas como evidência de que o *pipeline* adotado é adequado ao recorte municipal analisado e pode ser testado em outras unidades federativas.

Adicionalmente, realizou-se uma análise de robustez com o objetivo de examinar o impacto da inclusão dos valores projetados para o ano de 2023 no desempenho dos modelos. Para tal, os dois algoritmos de melhor desempenho (XGBoost e MLP) foram reestimados excluindo-se o referido ano da base de dados. Observou-se redução nas métricas preditivas (XGBoost: R^2 teste = 0,4424; R^2 CV = 0,4024; MLP: R^2 teste = 0,3823; R^2 CV = 0,3168), resultado consistente com a diminuição do tamanho amostral e da variabilidade estrutural disponível para estimação. Importante notar que, apesar da queda absoluta no desempenho, a hierarquia relativa entre os modelos foi preservada, mantendo-se o XGBoost como modelo de melhor desempenho relativo. Esses achados indicam que a inclusão de 2023 contribui para maior estabilidade estatística das estimativas, sem alterar o padrão comparativo entre os algoritmos avaliados.

5.4 Interpretação do Modelo Final (Importância das Variáveis)

Após a identificação do XGBoost como modelo de melhor desempenho preditivo, procedeu-se à sua interpretação por meio dos valores SHAP. Essa abordagem, fundamentada nos valores de *Shapley* da teoria dos jogos cooperativos, permite decompor cada predição em contribuições aditivas atribuídas às variáveis explicativas, garantindo consistência, localidade e aditividade nas explicações e atendendo à necessidade de transparência em aplicações educacionais baseadas em ML.

A Figura 3 apresenta o *beeswarm plo*t dos valores SHAP para o modelo XGBoost, sintetizando três dimensões fundamentais da explicação do modelo. A primeira é a magnitude do impacto, representada pela distância dos pontos em relação ao zero no eixo horizontal, indicando a força com que cada variável influencia a predição. A segunda refere-se à direção do efeito, sendo que valores SHAP positivos aumentam a estimativa do IDEB, enquanto valores negativos a reduzem. A terceira dimensão corresponde ao valor original de cada variável. Esses valores são codificados por uma escala de cores: pontos em rosa representam observações com valores altos da variável, enquanto pontos em azul representam valores baixos. Assim, a cor não indica impacto, mas sim o nível do preditor antes de entrar no modelo. Dessa forma, é possível verificar, por exemplo, se valores altos de um atributo tendem a produzir efeitos positivos (pontos rosados à direita) ou negativos (pontos rosados à esquerda) sobre a previsão do IDEB.

A visualização integrada dessas dimensões permite identificar as variáveis mais influentes na estrutura do modelo. Os resultados indicam que seis variáveis explicam de forma consistente o comportamento do IDEB municipal nos anos iniciais. A taxa de distorção idade-série apresentou o maior impacto absoluto entre todas as variáveis, refletido pela ampla dispersão dos valores SHAP. No *beeswarm plot* observa-se a concentração de pontos rosados à esquerda, indicando que valores elevados de distorção reduzem de forma consistente a previsão do IDEB. Esse padrão é coerente com evidências de que a taxa de distorção idade-série está associada à redução do desempenho acadêmico, o que explica sua relevância para o modelo.

O percentual de docentes sem formação superior (Grupo 5) também exerce influência negativa expressiva sobre o IDEB. Os pontos rosados posicionados à esquerda evidenciam que valores altos desse indicador estão associados à redução da previsão do IDEB municipal. A interpretação é compatível com o papel da formação docente na qualidade das práticas pedagógicas e no alinhamento curricular, aspectos amplamente discutidos na literatura sobre eficácia escolar.

Entre as variáveis com efeito positivo, destacam-se os repasses do Programa Criança Feliz, o valor da produção vegetal, o PIB per capita e a área plantada de lavouras temporárias e

permanentes. No caso do Programa Criança Feliz, valores mais elevados associaram-se a contribuições positivas para a previsão do IDEB, indicando possível influência indireta de políticas de apoio à primeira infância sobre resultados educacionais posteriores. O valor da produção vegetal e a área plantada atuam como indicadores de dinamismo econômico local, enquanto o PIB per capita representa capacidade econômica agregada. Em todos esses casos, valores mais elevados tendem a deslocar a predição para níveis superiores de IDEB, ainda que com magnitude inferior às variáveis diretamente educacionais.

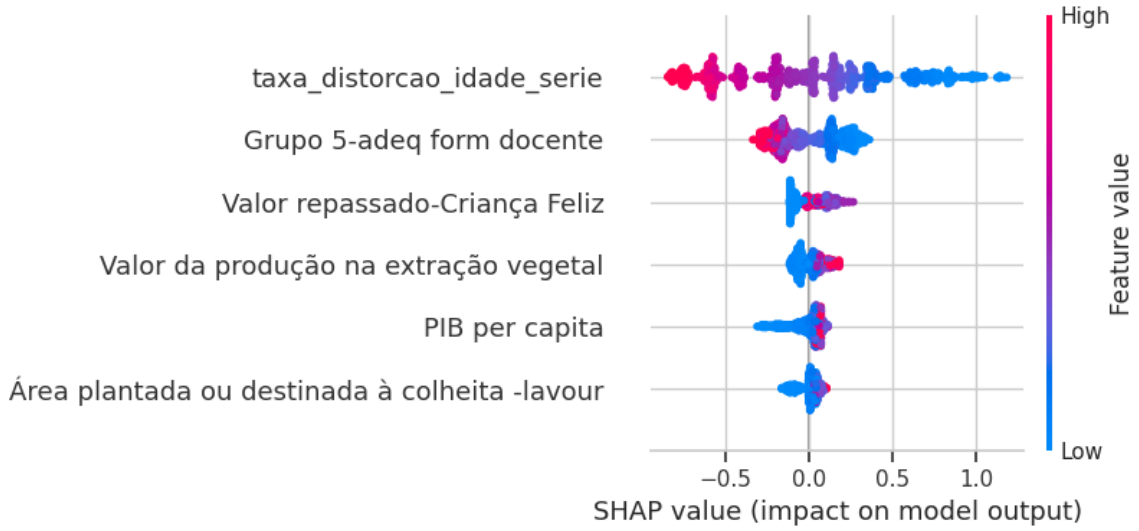

Figura 3: Importância global das variáveis segundo valores SHAP do modelo XGBoost.

Os resultados indicam que os fatores mais determinantes para explicar o desempenho municipal no IDEB dos anos iniciais são variáveis diretamente educacionais — notadamente a taxa de distorção idade-série e a qualificação docente. Variáveis econômicas e programas sociais atuam como moderadores positivos, contribuindo de maneira complementar para o desempenho escolar. Essa combinação reforça a consistência das estimativas do modelo e permite compreender como aspectos estruturais e pedagógicos se articulam na explicação das variações do IDEB entre os municípios piauienses.

Como complemento interpretativo, procedeu-se à análise de importância por permutação no modelo MLP, considerando que arquiteturas neurais não fornecem, de forma direta, medidas internas de importância estatisticamente interpretáveis. O método de permutação consiste em embaralhar aleatoriamente os valores de cada variável, mantendo-se as demais constantes, e mensurar o aumento no erro preditivo decorrente dessa perturbação. Quanto maior o aumento do erro (neste estudo, medido pelo MAE), maior a relevância da variável para o modelo.

A Figura 4 apresenta os resultados da importância por permutação aplicada ao MLP. Observa-se convergência substancial com os achados do SHAP no XGBoost: as mesmas seis variáveis figuram entre as mais influentes na predição do IDEB municipal. Essa convergência entre métodos distintos — um baseado em teoria dos jogos aplicado a modelo de árvore e outro baseado em perturbação aplicado a rede neural — reforça a consistência interpretativa dos resultados e reduz a probabilidade de que a importância observada seja artefato específico de um único algoritmo.

De forma integrada, os resultados indicam que variáveis diretamente associadas à estrutura educacional apresentam maior impacto na explicação do IDEB municipal, enquanto variáveis socioeconômicas atuam como condicionantes estruturais complementares. A consistência entre

modelos distintos fortalece a validade dos achados e contribui para maior transparência na interpretação dos mecanismos associados ao desempenho educacional no contexto analisado.

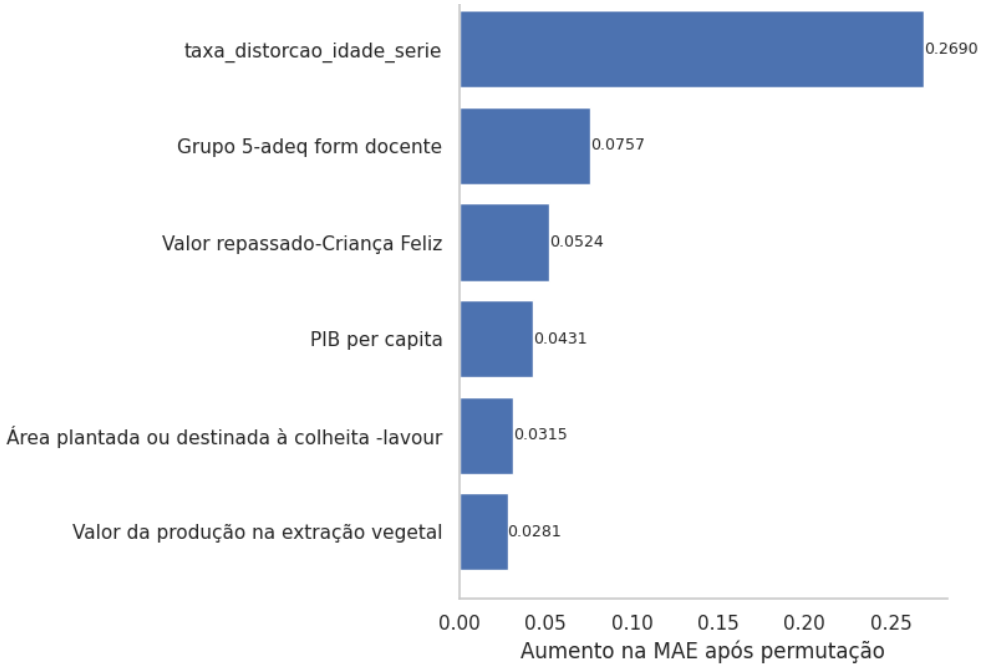

Figura 4: Importância das variáveis no modelo MLP obtida pelo método de permutação.

Do ponto de vista aplicado, esses resultados podem subsidiar o monitoramento educacional ao indicar dimensões prioritárias para investigação municipal, como taxa de distorção idade-série, formação docente e condições socioeconômicas locais. Ressalta-se, contudo, que as variáveis identificadas não devem ser interpretadas como efeitos causais diretos, mas como fatores associados às predições do modelo. Assim, o uso prático da abordagem deve ocorrer como apoio ao diagnóstico, à priorização de análises territoriais e à formulação de hipóteses de intervenção, sempre em conjunto com avaliações pedagógicas, institucionais e contextuais.

6 Conclusão

Este estudo avaliou 17 modelos preditivos, sendo dez algoritmos de ML e sete arquiteturas de DL, para estimar o IDEB dos anos iniciais do ensino fundamental nos municípios do estado do Piauí, a partir de variáveis educacionais, socioeconômicas e financeiras. A análise utilizou uma base multivariada com 1.262 registros e 126 variáveis, submetida a etapas de pré-processamento, seleção de atributos, validação cruzada e comparação estatística entre modelos.

Entre os modelos testados, o XGBoost apresentou o melhor desempenho médio no recorte analisado, com $R^2 = 0{,}5542$ na base de teste e $R^2 = 0{,}5455 \pm 0{,}0293$ em validação cruzada. Também obteve os menores erros médios e os melhores *rankings* relativos nas métricas R^2, MAE e RMSE. O teste de Friedman indicou diferença global significativa entre os modelos, mas as comparações *post-hoc* com correção de Holm não confirmaram diferenças estatisticamente significativas entre pares específicos. Assim, os resultados sustentam o melhor desempenho relativo do XGBoost no recorte analisado, mas não permitem afirmar superioridade estatística absoluta sobre todos os demais algoritmos.

A análise de importância das variáveis indicou que a taxa de distorção idade-série, a proporção de docentes sem formação superior e o PIB per capita municipal estiveram entre os fatores mais relevantes para explicar a variação do IDEB no estado do Piauí. Esses achados sugerem que o desempenho municipal piauiense nos anos iniciais está associado tanto a dimensões diretamente educacionais quanto a condições socioeconômicas locais, reforçando a importância de diagnósticos integrados para o acompanhamento educacional em escala municipal.

Entre as limitações do estudo, ressalta-se a dependência exclusiva de dados secundários públicos, que não contemplam variáveis subjetivas, como motivação e engajamento discente, nem aspectos qualitativos da prática pedagógica, da gestão escolar e do contexto familiar. Além disso, por se tratar de um estudo aplicado aos municípios do estado do Piauí, os resultados empíricos não devem ser generalizados diretamente para outras unidades federativas. A principal contribuição está na estruturação de uma abordagem metodológica replicável, baseada em dados públicos, seleção sistemática de variáveis, comparação de modelos, validação cruzada e interpretação dos preditores.

Para pesquisas futuras, encontra-se em andamento a ampliação do escopo territorial da abordagem para todos os municípios brasileiros, com coleta, tratamento, integração e padronização de bases educacionais, socioeconômicas e financeiras. A ampliação para a escala nacional tem como finalidade avaliar a estabilidade dos modelos, a variação regional dos fatores explicativos e a capacidade de generalização territorial do *pipeline* proposto. Também se encontra em planejamento e desenvolvimento inicial a construção de artefatos computacionais de simulação e visualização, destinados a traduzir os resultados preditivos em instrumentos exploratórios de apoio ao acompanhamento municipal. Nesse sentido, os resultados devem ser compreendidos como apoio ao diagnóstico e ao monitoramento educacional, e não como instrumento determinístico de decisão pública.

Referências

Benevento, M. A. (2024). O uso de algoritmos de aprendizagem de máquina para prever o desempenho do aluno (Tese de doutorado). Fundação Getulio Vargas, Escola de Administração de Empresas de São Paulo. Disponível em [link].

Brasil. (1988). Constituição da República Federativa do Brasil de 1988. Brasília, DF: Senado Federal. Disponível em [link].

Brasil. (1996). Lei n.º 9.394, de 20 de dezembro de 1996. Diário Oficial da União. Brasília, DF. Disponível em [link].

Brasil. (2020). Emenda constitucional n.º 108, de 26 de agosto de 2020. Diário Oficial da União. Brasília, DF. Disponível em [link].

Carreira, D., & Pinto, J. M. (2007). Custo aluno-qualidade inicial: Rumo à educação pública de qualidade no Brasil. São Paulo: Global; Campanha Nacional pelo Direito à Educação. Disponível em [link] [GS Search].

Chen, S., & Ding, Y. (2023). A machine learning approach to predicting academic performance in Pennsylvania's schools. Social Sciences, 12(3), 118. https://doi.org/10.3390/socsci12030118 [GS Search].

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Em Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). San Francisco: ACM. http://dx.doi.org/10.1145/2939672.2939785 [GS Search].

Conover, W. J. (1999). Practical nonparametric statistics (3rd ed.). New York: John Wiley & Sons. [GS Search].

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research, 7, 1–30. [GS Search].

Hanushek, E. A., & Woessmann, L. (2020). Education, knowledge capital, and economic growth. Em S. Sleeper (Ed.), The economics of education (2nd ed.). Amsterdam: Elsevier. https://doi.org/10.1016/B978-0-12-815391-8.00014-8 [GS Search].

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). New York: Springer. [GS Search].

Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial intelligence in education: Promises and implications for teaching and learning. (1st ed.). Center for Curriculum Redesign: Boston, MA, USA. [GS Search].

IBGE. (2023). Estatísticas sociais e econômicas. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística. Disponível em [link].

INEP. (2023). Resultados do IDEB e dos indicadores educacionais. Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Disponível em [link].

Keras Team. (2025). Keras API documentation. Disponível em [link].

Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the Boruta package. Journal of Statistical Software, 36(11), 1–13. https://doi.org/10.18637/jss.v036.i11 [GS Search].

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Em Advances in Neural Information Processing Systems (Vol. 30, pp. 4768–4777). https://doi.org/10.48550/arXiv.1705.07874 [GS Search].

Maia, J. S. Z., Bueno, A. P. A., & Sato, J. R. (2021). Assessing the educational performance of different Brazilian school cycles using data science methods. *PLOS ONE*, 16(3), e0248525. https://doi.org/10.1371/journal.pone.0248525 [GS Search].

MEC. (2023). SAGICAD: Sistema de Avaliação, Gestão da Informação e Cadastro Único. Brasília: Ministério da Educação. Disponível em [link].

Molnar, C. (2020). Interpretable machine learning: A guide for making black box models explainable (3rd ed.). Disponível em [link] [GS Search].

PNUD, IPEA, & FJP. (2022). Atlas do desenvolvimento humano no Brasil. Brasília: Programa das Nações Unidas para o Desenvolvimento, Instituto de Pesquisa Econômica Aplicada e Fundação João Pinheiro. Disponível em [link].

Rocha, F. A. F., Teixeira, J. C. M., & Melo, F. L. N. B. (2015). Análise dos fatores que influenciam o desempenho escolar dos alunos do ensino fundamental no estado do Rio Grande do Norte. Revista Interface, 12(1). [GS Search].

Rodrigues, L. S., Santos, M., Gomes, C. F. S., Choren, R., Goldschmidt, R., & Barbará, S. (2024). Transformers para previsão de desempenho acadêmico no ensino fundamental e médio. Revista Brasileira de Informática na Educação, 32, 213–241. https://doi.org/10.5753/rbie.2024.3661 [GS Search].

SICONFI. (2023). Sistema de Informações Contábeis e Fiscais do Setor Público Brasileiro. Brasília: Tesouro Nacional. Disponível em [link].

Silveira, A. A. D., Schneider, G., & Alves, T. (2023). Simulador de Custo-aluno Qualidade: Padrão de qualidade de referência, versão 02.2023. Curitiba; Goiânia: Laboratório de Dados Educacionais, UFPR; UFG. Disponível em [link].

SIOPE. (2023). Sistema de Informações sobre Orçamentos Públicos em Educação. Brasília: FNDE. Disponível em [link].

Soares, D. J. M., & Santos, W. (2024). Indicadores de avaliação de contexto e resultados educacionais no IDEB: Uma análise das escolas estaduais de ensino médio no Espírito Santo. Revista Brasileira de Estudos Pedagógicos, 105, e5872. https://doi.org/10.24109/2176-6681.rbep.105.5872 [GS Search].

Soares, J. F., & Araújo, R. J. (2006). Nível socioeconômico, qualidade e equidade das escolas de Belo Horizonte. Ensaio: Avaliação e Políticas Públicas em Educação, 14(50), 107–126. https://doi.org/10.1590/S0104-40362006000100008 [GS Search].

Souza, V. F., & Santos, T. C. B. (2021). Processo de mineração de dados educacionais aplicado na previsão do desempenho de alunos: Uma comparação entre técnicas de aprendizagem de máquina e aprendizagem profunda. Revista Brasileira de Informática na Educação, 29. https://doi.org/10.5753/RBIE.2021.29.0.519 [GS Search].

TensorFlow. (2025). TensorFlow documentation. Disponível em [link].

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B, 58(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x [GS Search].

UNESCO. (2022). Reimaginar juntos nossos futuros: Um novo contrato social para a educação. Paris: Organização das Nações Unidas para a Educação, a Ciência e a Cultura. Disponível em [link].

Wang, S., & Luo, B. (2024). Academic achievement prediction in higher education through interpretable modeling. PLOS ONE, 19(9), e0309838. https://doi.org/10.1371/journal.pone.0309838 [GS Search].

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education: Where are the educators?. International Journal of Educational Technology in Higher Education, 16, 39. https://doi.org/10.1186/s41239-019-0171-0 [GS Search].