Trabalho de Conclusão de Curso

Análise de Matrículas Escolares em Maceió: Um Estudo de Séries Temporais com LSTM

Artur Cavalcante de Jesus
acj@ic.ufal.br

Orientador:

Prof. Dr. Bruno Almeida Pimentel

Maceió, Fevereiro de 2025

Artur Cavalcante de Jesus

Análise de Matrículas Escolares em Maceió: Um Estudo de Séries Temporais com LSTM

Monografia apresentada como requisito parcial para obtenção do grau de Bacharel em Engenharia de Computação do Instituto de Computação da Universidade Federal de Alagoas.

Orientador:

Prof. Dr. Bruno Almeida Pimentel

Maceió, Fevereiro de 2025

Catalogação na Fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico

Bibliotecário: Marcelino de Carvalho Freitas Neto – CRB-4 - 1767

J58a Jesus, Artur Cavalcante de.

Análise de matrículas escolares em Maceió : um estudo de séries temporais com LSTM / Artur Cavalcante de Jesus. – 2025.

50 f. : il.

Orientador: Bruno Almeida Pimentel.

Monografia (Trabalho de conclusão de curso em Engenharia de Computação) - Universidade Federal de Alagoas, Instituto de Computação. Maceió, 2025.

Texto em inglês.

Bibliografia: f. 48-50.

1. Matrícula escolar. 2. Análise de séries temporais. 3. *Long Short-Term Memory*. 4. Inteligência artificial explicável. 5. Educação pública. I. Título.

CDU: 004.8

Monografia apresentada como requisito parcial para obtenção do grau de Bacharel em Engenharia de Computação do Instituto de Computação da Universidade Federal de Alagoas, aprovada pela comissão examinadora que abaixo assina.

Prof. Dr. Bruno Almeida Pimentel - Orientador
Instituto de Computação
Universidade Federal de Alagoas

Prof. Dr. Thiago Damasceno Cordeiro - Examinador
Instituto de Computação
Universidade Federal de Alagoas

Prof. Dr. Diego Carvalho do Nascimento - Examinador
Departamento de Matemática
Universidad de Atacama

Maceió, Fevereiro de 2025

Agradecimentos

Primeiramente, sou eternamente grato ao meu pai, Armando Santana de Jesus, à minha mãe, Talma Cristina Cavalcante Silva de Jesus, à minha irmã, Amanda Cavalcante de Jesus, e à minha namorada Manoella Andrade Bezerra Sampaio, pelo amor incondicional, apoio constante e encorajamento em cada etapa da minha vida. Seus valores, conselhos e dedicação foram fundamentais para que eu pudesse chegar até aqui.

Expresso minha sincera gratidão à Universidade Federal de Alagoas, à equipe administrativa do Instituto de Computação e aos docentes do curso de Engenharia de Computação, que me ofereceram um ambiente enriquecedor e diversas oportunidades para o meu desenvolvimento acadêmico e profissional.

Agradeço ao meu orientador, Prof. Dr. Bruno Almeida Pimentel, pelo apoio e orientação ao longo deste trabalho. Sua experiência, disponibilidade e confiança no meu potencial foram essenciais para o sucesso deste trabalho e para meu crescimento pessoal e acadêmico.

Por fim, expresso minha sincera gratidão aos amigos que tive a honra de conhecer ao longo desta jornada acadêmica. Em especial, agradeço a Guilherme de Oliveira Monteiro Peixoto, Pedro Antônio da Silva Pimentel Sousa e William Gabriel da Paz Rosendo, cuja amizade e apoio foram essenciais durante toda essa caminhada.

Resumo

A previsão da evolução das matrículas escolares é essencial para o planejamento educacional, possibilitando uma alocação eficiente de recursos e a formulação de políticas públicas fundamentadas em dados. Este estudo analisa os fatores que influenciam a quantidade de matrículas nos ensinos infantil, fundamental e médio na rede pública de Maceió, utilizando redes neurais *Long Short-Term Memory* (LSTM) para análise de séries temporais. A pesquisa foi conduzida com dados do Censo Escolar do INEP, abrangendo um período de dez anos. Além da análise preditiva, foram aplicadas técnicas de inteligência artificial explicável, como SHAP (*SHapley Additive Explanations*) e Causalidade de Granger, a fim de interpretar os fatores mais relevantes na variação das matrículas. Os resultados indicam que, no ensino infantil, a infraestrutura escolar, incluindo banheiros adaptados, equipamentos multimídia e parques infantis, influencia diretamente a decisão dos pais em matricular seus filhos. No ensino fundamental, a manutenção de um corpo docente adequado e a presença de bibliotecas bem estruturadas foram identificadas como fatores fundamentais para a retenção dos alunos. No ensino médio, a existência de laboratórios de ciências, infraestrutura esportiva e acesso à tecnologia demonstrou ser essencial para a permanência dos estudantes. Além disso, a análise revelou interdependências entre variáveis, sugerindo que cortes orçamentários em determinados setores podem gerar impactos negativos indiretos sobre a taxa de matrículas. Assim, os achados deste estudo fornecem subsídios para que gestores educacionais adotem estratégias baseadas em evidências, garantindo um planejamento mais eficiente e políticas educacionais voltadas à melhoria da qualidade da educação pública.

Palavras-chave: Matrícula escolar; Análise de séries temporais; *Long Short-Term Memory*; Inteligência artificial explicável; Educação pública.

Abstract

The prediction of school enrollment trends is essential for educational planning, enabling the efficient allocation of resources and the formulation of data-driven public policies. This study analyzes the factors influencing enrollment rates in preschool, elementary, and high school education in the public network of Maceió, using Long Short-Term Memory (LSTM) neural networks for time series analysis. The research was conducted with data from the INEP School Census, covering a ten-year period. In addition to predictive analysis, Explainable Artificial Intelligence techniques such as SHAP (SHapley Additive Explanations) and Granger Causality were applied to interpret the most relevant factors affecting enrollment trends. The results indicate that, in preschool education, school infrastructure, including adapted restrooms, multimedia equipment, and playgrounds, directly impacts parents' decisions regarding enrollment. In elementary school, maintaining an adequate teaching staff and providing well-structured libraries were identified as crucial factors for student retention. In high school, the availability of science laboratories, sports infrastructure, and access to technology proved to be essential for student retention. Furthermore, the analysis revealed interdependencies among variables, suggesting that budget cuts in specific areas may indirectly impact enrollment rates. Therefore, the findings of this study provide valuable insights for educational policymakers to adopt evidence-based strategies, ensuring more efficient planning and public policies aimed at improving the quality of public education.

Key-words: School enrollment; Time series analysis; Long Short-Term Memory; Explainable artificial intelligence; Public education.

Conteúdo

Lista de Figuras

Lista de Tabelas

1

Introdução

A gestão educacional eficiente exige um entendimento aprofundado dos fatores que influenciam a variação no número de matrículas ao longo dos anos. O Censo Escolar, conduzido pelo INEP, é a principal base de dados sobre a educação básica no Brasil, reunindo informações detalhadas sobre matrículas, infraestrutura escolar, disponibilidade de professores e características socioeconômicas dos alunos. A análise dessas variáveis permite compreender quais fatores impactam diretamente o comportamento das matrículas escolares, auxiliando na tomada de decisões estratégicas para o planejamento educacional.

Segundo [Lima e Sousa 2014], "os dados levantados subsidiam a elaboração de diagnóstico educacional do Brasil com objetivo de criar estratégias para acesso, permanência e sucesso dos alunos na escola". Isso demonstra que a análise do Censo Escolar não apenas permite acompanhar o número de matrículas ao longo do tempo, mas também oferece informações essenciais para identificar desigualdades regionais e propor ações corretivas para fortalecer o sistema educacional.

Os dados mais recentes do Censo Escolar 2023 revelam uma tendência preocupante de redução no número de matrículas na educação básica no Brasil. Entre 2022 e 2023, a rede pública registrou cerca de 77 mil matrículas a menos, totalizando 500 mil estudantes a menos em comparação com 2022. Em contrapartida, a rede privada teve um crescimento de 4,7% no número de alunos. O ensino fundamental apresentou um total de 26,1 milhões de estudantes matriculados, uma redução de 3% em relação a 2019, com os anos iniciais sofrendo uma queda mais acentuada (3,9%) do que os anos finais (1,9%). No ensino médio, foram registrados 7,7 milhões de estudantes, representando uma redução de 2,4% em relação ao ano anterior [EPSJV, Fiocruz 2024]. Esse cenário reforça a necessidade de analisar os fatores que afetam a permanência dos estudantes na escola e a distribuição de matrículas em diferentes regiões do país.

O município de Maceió representa um caso relevante dentro desse contexto nacional uma vez que enfrenta desafios significativos relacionados à evasão escolar e à regularidade das ma-

trículas. De acordo com o Painel dos Indicadores Educacionais, divulgado pelo INEP, a cidade registrou, em 2019, uma das maiores taxas de abandono escolar entre as capitais brasileiras, atingindo 2,1% no ensino fundamental da rede municipal [TATU 2021]. Além disso, um estudo do UNICEF aponta que, embora a taxa de abandono tenha diminuído nos últimos anos, ainda há disparidades regionais significativas identificadas [UNICEF 2020].

Outro fator relevante na dinâmica educacional de Maceió é a distorção idade-série, que afeta diretamente a progressão escolar e pode influenciar a continuidade dos estudos. Em 2016, 40,9% dos alunos do ensino fundamental estavam dois anos ou mais atrasados em relação à série ideal, totalizando cerca de 47 mil estudantes. Embora essa taxa tenha reduzido para 31,7% em 2019, o número ainda é expressivo e reforça a necessidade de análises detalhadas sobre os fatores que interferem na regularidade das matrículas escolares [UNICEF 2020].

Com base nesses desafios educacionais, este estudo tem como objetivo analisar as variáveis do Censo Escolar que impactam a quantidade de matrículas ao longo dos anos em Maceió. A identificação desses fatores pode fornecer subsídios valiosos para a gestão educacional, auxiliando na formulação de políticas públicas e na otimização da alocação de recursos, garantindo um planejamento mais eficiente para o setor educacional.

1.1 Justificativa

A previsão e o acompanhamento das matrículas escolares são fundamentais para o planejamento educacional e a gestão de recursos na educação pública. A análise das variáveis do Censo Escolar possibilita compreender quais fatores estruturais, pedagógicos e socioeconômicos influenciam diretamente a dinâmica das matrículas ao longo dos anos. Essas informações são essenciais para evitar superlotação de turmas, fechamento de escolas e a má distribuição de professores e infraestrutura.

Para [Souza e Oliveira 2012], diante dos dados do Censo Escolar, “são estabelecidas as políticas de correção dos desequilíbrios regionais e de promoção da equidade na oferta do ensino público”. Isso ressalta a relevância da análise dessas informações para identificar desigualdades educacionais e auxiliar na formulação de políticas públicas. Ao entender a influência das variáveis educacionais sobre as matrículas, torna-se possível adotar estratégias mais eficazes para reduzir distorções no acesso à educação e melhorar a qualidade do ensino.

A escolha de Maceió como objeto de estudo se justifica pelos desafios educacionais identificados nos últimos anos. Dados do INEP indicam que a cidade enfrenta altos índices de abandono escolar e distorção idade-série, refletindo a complexidade do cenário educacional [TATU 2021]. Embora esses problemas sejam significativos, o objetivo deste estudo não é focar exclusivamente na evasão, mas sim analisar como diferentes variáveis coletadas no Censo Escolar impactam o número de matrículas ao longo do tempo.

Para alcançar esse objetivo, este estudo emprega técnicas avançadas de inteligência arti-

ficial, especificamente redes neurais *Long Short-Term Memory* (LSTM), para modelar séries temporais e prever a evolução das matrículas. A abordagem baseada em LSTM permite capturar padrões históricos e dependências temporais entre os dados, proporcionando maior precisão na previsão de tendências futuras. Essa abordagem possibilita não apenas antecipar variações no número de matrículas, mas também identificar os fatores com maior impacto na dinâmica educacional, fornecendo subsídios estratégicos para a formulação de políticas públicas baseadas em dados.

Ao compreender os principais fatores que influenciam essa dinâmica, espera-se contribuir para a melhoria da distribuição de recursos, estruturação do ensino e planejamento estratégico da educação municipal. Dessa forma, o estudo não apenas impulsiona o uso de inteligência artificial na análise educacional, mas também pode ser utilizado como referência para estudos semelhantes em outras regiões do país.

1.2 Objetivo Geral

Analisar o impacto de variáveis educacionais no número de matrículas escolares no munícipio de Maceió na rede pública nos ensinos infantil, fundamental e médio, utilizando modelos de séries temporais (LSTM) e ferramentas de explicabilidade de IA.

1.3 Objetivos Específicos

- Consolidar e processar os dados educacionais do Censo Escolar de 2013 a 2022 filtrando apenas o município de Maceió.
- Aplicar modelos de aprendizado profundo baseados em LSTM para prever o número de matrículas nos ensinos infantil, fundamental e médio.
- Utilizar ferramentas de IA Explicável para interpretar os fatores mais importantes que influenciam o número de matrículas.
- Propor recomendações com base nos resultados obtidos para orientar gestores educacionais.

1.4 Estrutura

Com essa abordagem, este trabalho organiza-se em cinco capítulos. O Capítulo 1 apresenta a introdução, motivação e objetivos do estudo. No Capítulo 2, revisamos a literatura sobre aprendizado de máquina, interpretabilidade de modelos e estudos relacionados à análise de dados educacionais. O Capítulo 3 descreve a metodologia adotada e o processamento dos dados. O

Capítulo 4 traz os resultados e discussões, enquanto o Capítulo 5 conclui o trabalho, destacando as contribuições e sugestões para estudos futuros.

2

Fundamentação

Este capítulo apresenta os conceitos teóricos fundamentais para o desenvolvimento deste trabalho. Inicialmente, são discutidos aspectos relacionados às séries temporais, incluindo sua definição e principais características. Em seguida, aborda-se o papel do *Machine Learning* na modelagem de séries temporais, com destaque para as Redes Neurais Recorrentes (RNNs) e Redes LSTM (*Long Short-Term Memory*). Por fim, exploram-se as métricas de desempenho utilizadas para avaliar a acurácia dos modelos e os conceitos relacionados à *Explainable AI*, incluindo SHAP e Causalidade de Granger.

2.1 Séries Temporais

Uma série temporal é um conjunto de observações ordenadas no tempo, não necessariamente igualmente espaçadas, que apresentam dependência serial, isto é, dependência entre instantes de tempo. A notação usada aqui para denotar uma série temporal é $S_1, S_2, S_3 ..., S_T$ que indica uma série de tamanho T. Uma grande quantidade de fenômenos de natureza física, biológica, econômica, etc. pode ser enquadrada nesta categoria. A maneira tradicional de analisar uma série temporal é através da sua decomposição nas componentes de tendência, ciclo e sazonalidade. [Morettin 1987].

A **tendência** de uma série indica o seu comportamento "de longo prazo", isto é, se ela cresce, decresce ou permanece estável, e qual a velocidade destas mudanças. Nos casos mais comuns trabalha-se com tendência constante, linear ou quadrática, como ilustrado na Figura 2.1.

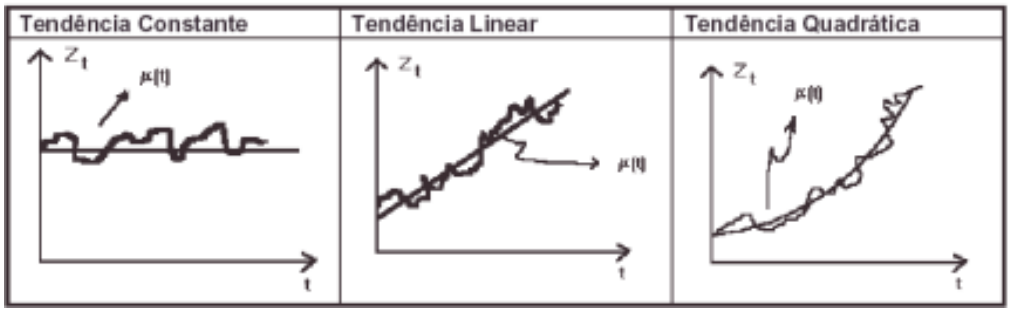

Figura 2.1: Tendências de uma série temporal [Barros 2003].

Os **ciclos** são caracterizados pelas oscilações de subida e de queda nas séries, de forma suave e repetida, ao longo da componente de tendência. Por exemplo, ciclos relacionados à atividade econômica ou ciclos meteorológicos.

A **sazonalidade** em uma série corresponde às oscilações de subida e de queda que sempre ocorrem em um determinado período do ano, do mês, da semana ou do dia. A diferença essencial entre as componentes sazonal e cíclica é que a primeira possui movimentos facilmente previsíveis, ocorrendo em intervalos regulares de tempo, enquanto que movimentos cíclicos tendem a ser irregulares.

O **ruído** (ou componente aleatório) representa as variações imprevisíveis e não sistemáticas presentes nos dados. Essas flutuações são causadas por fatores não modelados, erros de medição ou aleatoriedades intrínsecas ao fenômeno estudado. Diferentemente dos demais componentes, o ruído não segue um padrão reconhecível e é geralmente tratado como um resíduo após a remoção da tendência, ciclo e sazonalidade.

Em geral ao estudarmos uma série temporal estamos interessados em:

a) Análise e modelagem da série temporal - descrever a série, verificar suas características mais relevantes e suas possíveis relações com outras séries;

b) Previsão na série temporal - a partir de valores históricos da série (e possivelmente de outras séries também) procura-se estimar previsões de curto prazo (*forecast*). O número de instantes à frente para o qual é feita a previsão é chamado de horizonte de previsão.

2.2 *Machine Learning*

O *Machine Learning* (Aprendizado de Máquina) é uma subárea da Inteligência Artificial que desenvolve algoritmos capazes de aprender padrões complexos a partir de dados, sem serem explicitamente programados. Esses algoritmos podem ser categorizados em três abordagens principais: aprendizado supervisionado, aprendizado não supervisionado e aprendizado por reforço.

- **Aprendizado supervisionado**: No aprendizado supervisionado, o modelo é treinado utilizando um conjunto de dados rotulados, ou seja, dados que já possuem uma classificação

ou resultado esperado. Essa abordagem permite que o modelo aprenda a mapear entradas para saídas específicas, tornando-o capaz de realizar previsões precisas com base em novos dados de entrada [Bochie et al. 2020]. É amplamente utilizada em cenários onde há uma grande quantidade de dados disponíveis e os valores de saída do processo são conhecidos.

- **Aprendizado não supervisionado**: Essa abordagem é aplicada quando não há um conjunto de dados rotulados disponível. Nesse contexto, o objetivo principal é identificar padrões ou estruturas ocultas nos dados, o que pode ser útil em diversas situações, como detecção de anomalias ou agrupamento de dados semelhantes [Goodfellow e Bengio 2016]. Essa técnica é particularmente adequada para problemas em que a saída esperada não é clara ou quando o conjunto de dados é muito extenso para ser rotulado manualmente.

- **Aprendizado por reforço**: No aprendizado por reforço, o modelo aprende por meio de interações com um ambiente dinâmico, recebendo recompensas ou penalidades com base nas ações que realiza. O objetivo é que o modelo maximize a recompensa acumulada ao longo do tempo, aprendendo a tomar decisões que otimizem o desempenho em uma determinada tarefa. Essa abordagem é amplamente utilizada em áreas como robótica, jogos e controle de sistemas, onde a tomada de decisões sequenciais é crucial.

Dentre as técnicas avançadas de *Machine Learning*, destaca-se o *Deep Learning* (Aprendizado Profundo), que busca desenvolver modelos capazes de interpretar e aprender a partir de dados complexos e de alta dimensionalidades. Esses modelos são construídos com redes neurais profundas, projetadas para representar os dados de forma hierárquica, extraindo características progressivamente mais abstratas em suas camadas mais profundas [Bochie et al. 2020]. Essas redes são especialmente eficazes em tarefas que envolvem dados complexos, como imagens, áudios ou sequências temporais.

As redes neurais artificiais profundas são formadas por múltiplas camadas de neurônios, que funcionam como unidades de processamento. Esses neurônios recebem entradas, realizam cálculos matemáticos e produzem saídas. Cada camada da rede aplica uma transformação não-linear aos dados, permitindo que o modelo aprenda a identificar características cada vez mais complexas [Goodfellow e Bengio 2016]. A camada inicial recebe os dados brutos, como imagens ou áudios, enquanto as camadas subsequentes refinam essas informações, gerando representações mais sofisticadas.

Nesse contexto, dentre as técnicas de *Deep Learning*, as *Recurrent Neural Networks* (RNNs) e suas variantes, como as *Long Short-Term Memory* (LSTM), têm se destacado por sua capacidade de lidar com dados sequenciais, como séries temporais. Tais arquiteturas permitem a modelagem de padrões dinâmicos ao longo do tempo e têm aplicações práticas em áreas como previsão de demanda, análise financeira e previsão de matrículas educacionais. A seguir, exploraremos essas arquiteturas em detalhes.

2.2.1 *Recurrent Neural Networks* - RNNs

As Redes Neurais Recorrentes (RNNs – *Recurrent Neural Networks*) foram desenvolvidas para modelar relações temporais em dados sequenciais, permitindo que informações de estados anteriores influenciem previsões futuras. Diferente das redes neurais tradicionais, que assumem independência entre as entradas, as RNNs mantêm uma memória interna por meio de conexões recorrentes, tornando-as ideais para tarefas como reconhecimento de fala, processamento de texto e previsão de séries temporais [Christoper 2015].

Uma das principais características dessas redes é sua capacidade de armazenar e processar informações ao longo do tempo, utilizando mecanismos que ajustam a influência de estados passados sobre o estado atual. Isso é possível através da atualização periódica dos pesos e viés da rede, permitindo uma adaptação contínua aos dados [Bakhtierzhon e Petrusevich 2024]).

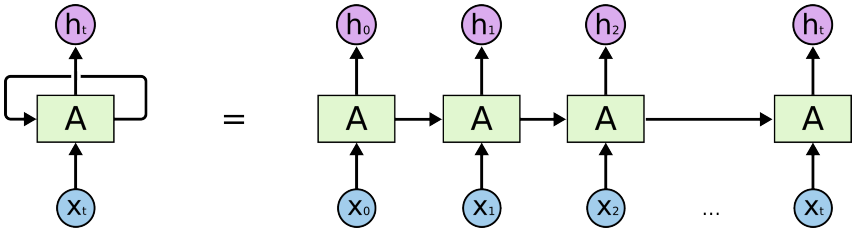

Figura 2.2: RNN desmembrada (Christopher, 2015).

No entanto, RNNs tradicionais enfrentam dificuldades ao capturar dependências de longo prazo, pois sofrem com desaparecimento do gradiente (*vanishing gradient*) e explosão do gradiente (*exploding gradient*), o que prejudica o treinamento e a eficiência da rede. Esses desafios foram amplamente discutidos por Bengio e Simard [Bengio e Simard 2019] e também por [Josef 1991], destacando a limitação das RNNs na aprendizagem de padrões temporais distantes.

O problema do desaparecimento do gradiente ocorre quando uma rede neural recorrente não consegue transmitir informações úteis do gradiente da saída de volta para as entradas iniciais, prejudicando o aprendizado. Por outro lado, a explosão do gradiente acontece quando os gradientes de erro se acumulam excessivamente, resultando em grandes atualizações nos pesos durante o treinamento. Esses fenômenos tornam os modelos instáveis e dificultam o aprendizado adequado a partir dos dados de treinamento, comprometendo sua eficiência [Brownlee 2019].

O gradiente descendente (*gradient descent*) é um método de otimização amplamente utilizado para encontrar o mínimo local de uma função de perda (*loss function*). Em algoritmos de aprendizado de máquina, o processo geralmente começa em um ponto inicial, onde o gradiente da função de perda é calculado para orientar a minimização do erro. De forma simplificada, o cálculo do gradiente envolve a derivada da função de perda em um ponto específico, indicando a direção de maior crescimento da função. A partir dessa informação, é possível ajustar os parâmetros do modelo para reduzir gradualmente o erro [Kirsten 2019].

Considere o caso clássico do mercado de ações, onde é necessário prever o preço de um ativo com base em eventos que podem ocorrer de forma esporádica, como crises econômicas ou lançamentos de novos produtos. Em situações como essa, é fundamental que o modelo seja capaz de considerar um contexto abrangente, incluindo informações históricas distantes. No entanto, à medida que essas lacunas temporais aumentam, as Redes Neurais Recorrentes (RNNs) enfrentam dificuldades em aprender a conectar essas informações de forma eficiente, tornando-se incapazes de capturar relações de longo prazo, conforme ilustrado na Figura 2.3.

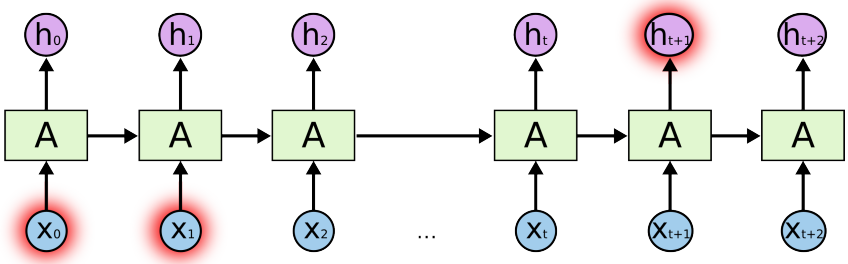

Figura 2.3: RNNs com dependências de longo prazo (Christopher, 2015)

Após identificar as limitações das Redes Neurais Recorrentes (RNNs), como o desaparecimento do gradiente (*vanishing gradient*) e a explosão do gradiente (*exploding gradient*), [Hochreiter e Schmidhuber 1997] propuseram a arquitetura *Long Short-Term Memory* (LSTM). Esse modelo aprimorado introduz portas de entrada, esquecimento e saída, que regulam o fluxo de informações, permitindo a retenção seletiva de estados relevantes ao longo do tempo. Dessa forma, as LSTMs conseguem preservar informações importantes mesmo em sequências de longo prazo, superando as dificuldades das RNNs tradicionais.

Estudos recentes apontam que as LSTMs apresentam desempenho superior na modelagem de séries temporais, principalmente quando comparadas a métodos estatísticos tradicionais, como ARIMA e ETS. Isso se deve à sua capacidade de capturar padrões temporais complexos e minimizar a perda de informações ao longo do processo [Bakhtierzhon e Petrusevich 2024].

2.2.2 Redes LSTM (*Long Short-Term Memory*)

As LSTMs foram propostas para superar as limitações das RNNs tradicionais ao capturar dependências mais extensas — algo fundamental em séries temporais que apresentam tendências de longo prazo ou picos sazonais. Em vez de permitir que informações antigas se percam rapidamente, as LSTMs incluem mecanismos internos capazes de selecionar quais dados devem ser mantidos ou descartados em cada passo de tempo, garantindo que "memórias" relevantes permaneçam disponíveis mesmo em sequências longas. A Figura 2.4 ilustra detalhadamente essa estrutura.

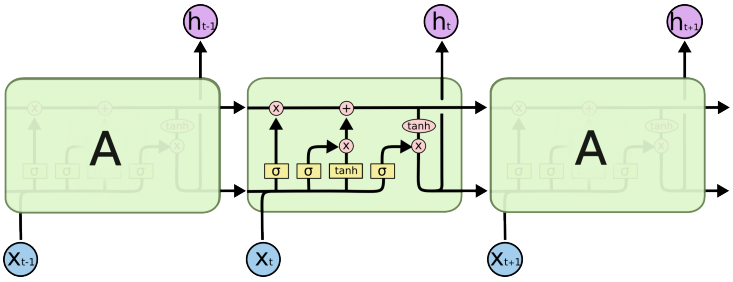

Figura 2.4: Unidade LSTM (Christopher, 2015).

O comportamento da memória interna é regulado por estruturas chamadas de portões ou (*gates*), em inglês. Eles são compostos de uma função de ativação sigmoide e uma operação de multiplicação por pontos, conforme ilustrado na Figura 2.5.

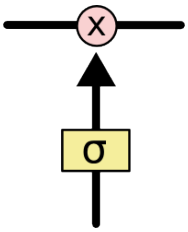

Figura 2.5: Ilustração de um *gate*. (Christopher, 2015).

A primeira função realizada por uma unidade LSTM é decidir quais informações do estado anterior serão descartadas e quais serão mantidas. Esse papel é desempenhado pela *forget gate*, que avalia a saída do estado anterior (h_{t-1}) em conjunto com a entrada atual (x_t). A partir dessa análise, a porta de esquecimento gera um vetor de valores (f_t) que varia entre 0 e 1 para cada elemento da memória anterior (c_{t-1}). Um valor próximo de 0 indica que a informação será descartada, enquanto valores próximos de 1 permitem que a informação seja preservada e utilizada no estado atual [Lv et al. 2022].

A operação realizada pela *forget gate* é representada pela Equação 2.1:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{2.1}$$

Nessa equação (W_f) corresponde à matriz de pesos que relaciona (h_{t-1}) e (x_t), enquanto (b_f) é o vetor de bias que ajusta as ativações resultantes. Já (σ) indica a função de ativação sigmoide, cujo intervalo de saída varia entre 0 e 1, permitindo filtrar de forma gradativa as informações não essenciais.

Essa etapa é essencial para que a unidade LSTM processe dados sequenciais de forma eficiente, filtrando informações irrelevantes ou desnecessárias e mantendo apenas os elementos essenciais para compor o contexto da memória no estado atual.

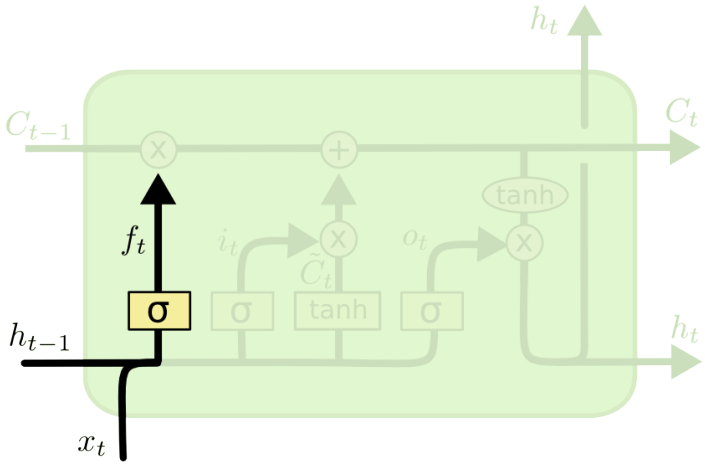

Figura 2.6: Ilustração de um *forget gate*. (Christopher, 2015).

O próximo passo no funcionamento da unidade LSTM é determinar quais novas informações serão incorporadas ao estado da célula. Esse processo é controlado pela *input gate* (i_t), que utiliza uma função de ativação sigmoide para avaliar e selecionar os valores que devem ser atualizados [Ni et al. 2020]. Essa avaliação é descrita matematicamente pela Equação 2.2:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2.2}$$

Nessa equação (W_i) corresponde à matriz de pesos que relaciona (h_{t-1}) e (x_t), enquanto (b_i) responsável pelo ajuste fino da ativação resultante. A função de ativação (σ) representa a sigmoide.

Em seguida, uma função tangente hiperbólica (tanh) gera um vetor de novos valores candidatos ($\tilde{C}_t$), que representa informações potenciais a serem adicionadas ao estado atual da célula. Esse cálculo pode ser expresso pela Equação 2.3:

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{2.3}$$

Nessa expressão (W_c) corresponde à matriz de pesos que relaciona (h_{t-1}) e (x_t), enquanto (b_c) responsável pelo ajuste fino da ativação resultante. A função de ativação (σ) representa a sigmoide. Já a tangente hiperbólica (tanh) produz valores na faixa de -1 a 1, permitindo que o modelo represente tanto variações negativas quanto positivas na informação a ser incorporada ao estado atual da célula.

A combinação desses dois componentes — i_t e $\tilde{C}_t$ — resulta na atualização do estado da célula, permitindo que apenas as informações mais relevantes sejam incorporadas. Esse mecanismo garante que o modelo armazene novos dados enquanto substitui aqueles que já não são necessários.

A Figura 2.7 ilustra o funcionamento da *input gate*, destacando como as entradas (h_{t-1}) e (x_t) são processadas para gerar os valores de atualização (i_t) e os novos candidatos ($\tilde{C}_t$). Esse

processo é fundamental para o aprendizado eficiente em dados sequenciais.

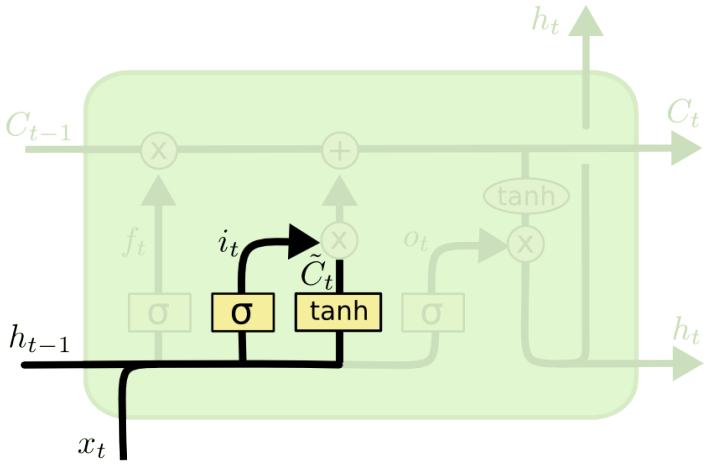

Figura 2.7: Ilustração de um *input gate*. (Christopher, 2015).

Após decidir quais informações serão mantidas e quais serão atualizadas, o próximo passo é calcular o novo estado da célula (C_t). Esse processo combina informações do estado anterior (C_{t-1}) com os novos valores gerados nas etapas anteriores.

Primeiro, o estado anterior é multiplicado pelo vetor da *forget gate* (f_t), descartando as informações consideradas irrelevantes. Em seguida, os valores candidatos gerados pela função tangente hiperbólica ($\tilde{C}_t$) são ponderados pelo vetor da *input gate* (i_t), adicionando ao estado atual apenas as informações relevantes. Esse cálculo é descrito pela Equação 2.4:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{2.4}$$

Nessa fórmula, o operador ($*$) representa a multiplicação elemento por elemento, garantindo que cada valor seja ajustado de maneira controlada.

Esse mecanismo assegura que o modelo consiga manter informações essenciais do passado enquanto incorpora novas informações ao estado da célula. A Figura 2.8 ilustra esse processo, destacando como os vetores f_t, C_{t-1}, i_t e $\tilde{C}_t$ interagem para produzir o novo estado C_t. Essa atualização é fundamental para que a unidade LSTM capture dependências temporais de forma eficiente e robusta.

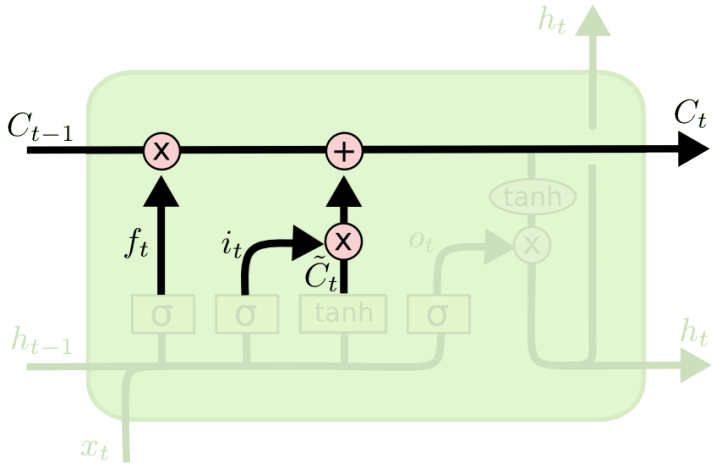

Figura 2.8: Processo de atualização do estado da célula (Christopher, 2015).

O último passo na unidade LSTM é determinar a saída (h_t) que será enviada para o próximo estado da rede. Essa saída é baseada no estado atual da célula (C_t), mas em uma versão filtrada e transformada.

Primeiramente, a *output gate* (o_t) avalia quais partes do estado da célula serão utilizadas na saída. Essa decisão é realizada por meio de uma função de ativação sigmoide, que processa as entradas h_{t-1} e x_t, conforme a seguinte Equação 2.5:

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{2.5}$$

Nessa expressão (W_o) corresponde à matriz de pesos que relaciona (h_{t-1}) e (x_t), enquanto (b_o) responsável pelo ajuste fino da ativação resultante. A função de ativação (σ) representa a sigmoide gera valores no intervalo de 0 a 1, permitindo determinar o grau de influência que a memória da célula exercerá sobre a saída final.

Em seguida, o estado da célula (C_t) passa por uma função tangente hiperbólica (tanh), que normaliza os valores para o intervalo entre -1 e 1. A saída final é obtida multiplicando-se o resultado da tangente hiperbólica pelo vetor (o_t), resultando na seguinte Equação 2.6:

$$h_t = o_t * \tanh(C_t) \tag{2.6}$$

Esse processo garante que a unidade LSTM produza uma saída que reflete tanto o estado interno atual quanto os ajustes definidos pela porta de saída, permitindo que informações essenciais sejam propagadas ao próximo estado da rede.

A Figura 2.9 ilustra esse mecanismo, destacando como as entradas h_{t-1}, x_t, C_t, e as portas de saída interagem para gerar a saída h_t. Esse passo é crucial para garantir que a rede capture e transmita as informações mais relevantes de forma eficiente em tarefas sequenciais.

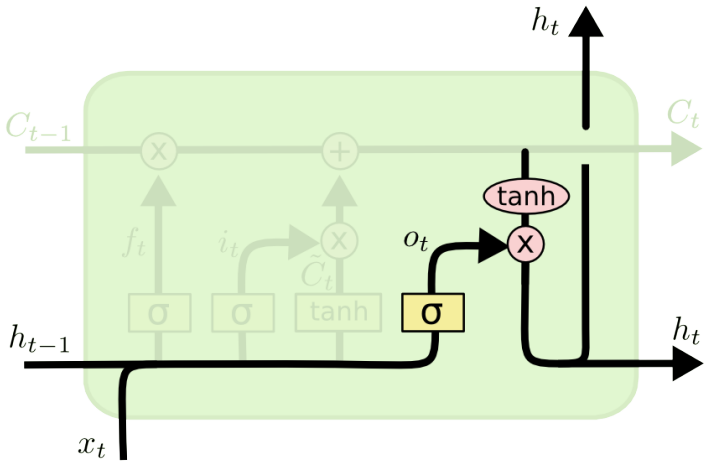

Figura 2.9: Ilustração de um *output gate*. (Christopher, 2015).

2.3 Métricas de Desempenho

Para verificar o desempenho de Redes Neurais em problemas de séries temporais é comum utilizar métricas que comparem os valores preditos com valores reais. Após uma pesquisa na literatura foi possível entender que existem algumas métricas que são bastantes utilizadas, então para acompanhar a literatura foram utilizadas as métricas:

2.3.1 *Mean Squared Error* – MSE

O erro quadrático médio (MSE) é uma medida comumente utilizada para avaliar a acurácia de modelos de aprendizado de máquina. O MSE é calculado elevando ao quadrado cada erro individual (diferença entre o valor real Y_i e o valor previsto $\hat{Y}_i$) e, em seguida, calculando a média desses erros quadráticos para todas as N observações do conjunto de dados. Matematicamente, o MSE é definido pela Equação 2.7:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \hat{Y}_i\right)^2 \tag{2.7}$$

Ao elevar ao quadrado cada erro, o MSE atribui maior peso a erros maiores, o que pode ser útil para penalizar discrepâncias significativas. No entanto, essa característica torna o MSE sensível a valores extremos (*outliers*), limitando sua aplicação em cenários com alta variabilidade nos dados [Géron 2019].

2.3.2 *Root Mean Squared Error* – RMSE

O erro quadrático médio da raiz (RMSE) é uma métrica derivada do MSE, amplamente utilizada em problemas de regressão e séries temporais. O RMSE é obtido calculando a raiz quadrada

da média dos erros quadráticos entre os valores reais Y_i e os previstos $\hat{Y}_i$, para um total de N observações. Sua principal vantagem é expressar o erro na mesma escala dos dados originais, facilitando a interpretação. A Equação 2.8 abaixo define o RMSE:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \hat{Y}_i\right)^2} \tag{2.8}$$

Assim como o MSE, o RMSE penaliza erros grandes de forma acentuada, o que o torna igualmente sensível a *outliers* [Géron 2019]. Por isso, embora seja uma métrica valiosa para avaliar desempenho global, seu uso deve considerar a natureza dos dados e os objetivos específicos do modelo.

2.3.3 *Mean Absolute Error* – MAE

O erro médio absoluto (MAE) é uma medida de avaliação utilizada em modelos de aprendizado de máquina, calculado a partir da média dos erros absolutos entre os valores reais Y_i e os valores previstos $\hat{Y}_i$, onde Y_i representa o valor real da i-ésima observação, $\hat{Y}_i$ é o valor previsto pelo modelo para a mesma observação, e N corresponde ao número total de observações no conjunto de dados. Ao utilizar o módulo de cada erro, o MAE evita que erros positivos e negativos se cancelem, tornando-o menos sensível a valores extremos (*outliers*) em comparação com métricas como o MSE [Géron 2019]. A Equação 2.9 apresenta o cálculo do MAE:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|Y_i - \hat{Y}_i| \tag{2.9}$$

2.3.4 Coeficiente de determinação - R^2

O coeficiente de determinação (R^2) é uma métrica utilizada em modelos de regressão para avaliar a proporção da variância da variável dependente Y que é explicada **linearmente** pelo modelo. Valores de R^2 variam entre 0 e 1, sendo que valores próximos de 1 indicam que o modelo explica grande parte da variabilidade dos dados por meio da relação linear, enquanto valores próximos de 0 sugerem que o modelo não captura adequadamente a relação linear entre as variáveis. O R^2 é calculado conforme a Equação 2.10, comparando a soma dos quadrados dos resíduos (SSR), que corresponde a $\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2$, com a soma total dos quadrados (SST), definida como $\sum_{i=1}^{N}(Y_i - \bar{Y})^2$, onde $\bar{Y}$ é a média dos valores reais Y_i:

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{N}(Y_i - \bar{Y})^2} \tag{2.10}$$

2.4 *Explainable AI*

A Inteligência Artificial Explicável (*Explainable AI* - XAI) é um campo emergente da inteligência artificial que busca reduzir o *gap* entre a complexidade dos modelos de aprendizado de máquina e a compreensão humana. Com o aumento do uso de modelos de *machine learning* em decisões críticas, como saúde, finanças e justiça, a transparência e interpretabilidade tornaram-se essenciais para aumentar a confiança nos sistemas automatizados. Modelos baseados em aprendizado profundo (*deep learning*) são frequentemente considerados "caixas-pretas" devido à dificuldade de explicar suas decisões, o que pode gerar preocupações éticas e viés algorítmico em aplicações sensíveis [Velden 2024].

Atualmente, os principais métodos de XAI podem ser classificados em três categorias:

- Explicações visuais, como *heatmaps* e mapas de saliência, que destacam as áreas de uma imagem utilizadas na tomada de decisão;
- Explicações textuais, que fornecem descrições interpretáveis sobre o funcionamento do modelo;
- Explicações baseadas em exemplos, que usam casos similares para justificar as predições da IA.

Entre essas abordagens, as explicações visuais são amplamente utilizadas, especialmente na área médica, onde modelos XAI ajudam a identificar padrões em imagens de diagnóstico médico [Velden 2024].

A XAI enfrenta um desafio conhecido como *trade-off* entre interpretabilidade e desempenho:

- Modelos simples, como regressões lineares, são altamente interpretáveis, mas muitas vezes incapazes de capturar relações complexas nos dados;
- Modelos mais sofisticados, como redes neurais profundas, oferecem alta precisão, mas funcionam como "caixas-pretas", tornando difícil entender suas decisões [Velden 2024]

Para mitigar esse problema, pesquisas recentes sugerem o uso de abordagens explicáveis por design (*explainable-by-design*), que incorporam a interpretabilidade desde a concepção do modelo. Além disso, novas direções para XAI incluem técnicas de IA causal, que buscam explicar não apenas correlações, mas também relações de causa e efeito nos modelos de aprendizado de máquina [Velden 2024].

Dentre as técnicas mais utilizadas na XAI, duas abordagens serão detalhadas neste trabalho: o **SHAP (*SHapley Additive Explanations*)** e o uso de **Causalidade de Granger**, que permitem analisar, respectivamente, os impactos das variáveis nas predições e as relações causais entre as variáveis no contexto temporal.

2.4.1 SHAP - *SHapley Additive Explanations*

O SHAP, baseado na teoria dos jogos, é uma das ferramentas mais poderosas na XAI. Ele utiliza os valores de Shapley, que fornecem uma maneira matemática de calcular as contribuições individuais de cada variável na predição de um modelo, sendo útil tanto em explicações locais quanto globais [Lundberg e Lee 2017].

No nível local, o SHAP calcula como cada variável contribuiu para a predição de uma instância específica. O gráfico de força é amplamente utilizado para destacar essas contribuições, onde a combinação de fatores positivos e negativos resulta no valor predito pelo modelo.

Figura 2.10: Resultado do SHAP Local.

No nível global, o SHAP gera gráficos de resumo que mostram o impacto médio das variáveis no modelo. Esses gráficos permitem identificar quais variáveis mais influenciam o desempenho do modelo e de que forma elas afetam a saída, considerando valores baixos e altos para cada variável.

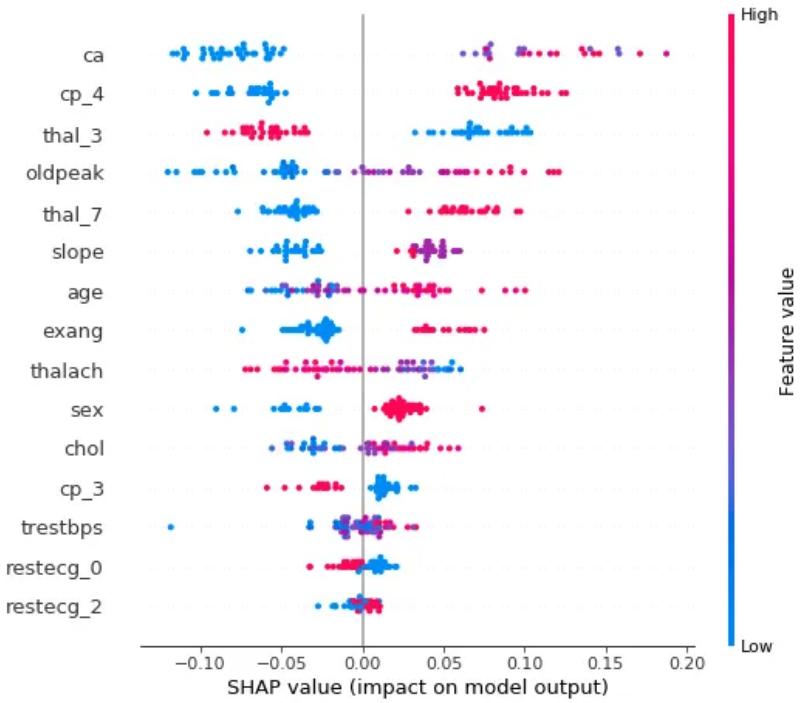

Figura 2.11: Resultado do SHAP Global.

Os valores de Shapley são calculados conforme a Equação 2.11:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} \cdot [v(S \cup \{i\}) - v(S)] \tag{2.11}$$

Onde:
- ϕ_i: valor de Shapley para a variável i;
- S: subconjunto de variáveis sem i;
- N: conjunto completo de variáveis;
- $v(S)$: função de predição considerando o subconjunto S.

2.4.2 Causalidade de Granger

A Causalidade de Granger é um método estatístico amplamente utilizado para determinar a relação causal entre variáveis temporais, especialmente em análises de séries temporais multivariadas. Proposta por Clive Granger [Granger 1969], essa abordagem testa se o histórico de uma variável pode melhorar as previsões de outra variável, estabelecendo uma relação de causa e efeito no contexto temporal. No entanto, é importante destacar que a Causalidade de Granger não implica causalidade verdadeira no sentido filosófico, mas sim uma capacidade preditiva fundamentada em correlações temporais.

Uma variável X_t é considerada causal no sentido de Granger para uma variável Y_t se os valores passados de X_t melhorarem a previsão de Y_t, além do que seria possível apenas com os dados históricos de Y_t. O teste utiliza modelos autorregressivos, como o VAR (Modelos Autorregressivos Vetoriais), para avaliar se os *lags* de X_t são estatisticamente significativos ao prever Y_t.

A ideia por trás da Causalidade de Granger (para séries temporais univariadas) é considerar o modelo descrito pela Equação 2.12:

$$Y_t = \beta_0 + \sum_{i=1}^{k} \beta_i Y_{t-i} + \sum_{j=1}^{m} \alpha_j X_{t-j} + \varepsilon_t \tag{2.12}$$

Onde:
- Y_t é a variável dependente;
- X_t é a variável independente testada como causal;
- k é o número de defasagens (*lags*) para Y_t;
- m é o número de defasagens (*lags*) para X_t;
- β_0 é a constante do modelo;
- β_i e α_j são os coeficientes a serem estimados;
- ε_t é o termo de erro ou ruído branco.

Dizemos que X_t Granger-causa Y_t se os valores passados de X_t ajudam a prever o valor presente de Y_t. Para testar se X_t Granger-causa Y_t, considera-se o seguinte teste:

$$H_0 : \alpha_1 = \cdots = \alpha_m = 0 \quad \text{vs.} \quad H_1 : \alpha_s \neq 0, \text{ para pelo menos um } s \in \{1, \cdots, m\}.$$

No teste acima, com a rejeição da hipótese nula, pode-se concluir que X_t Granger-causa Y_t. Semelhantemente, considerando o modelo dado pela Equação 2.13:

$$X_t = \beta_0 + \sum_{i=1}^{k} \beta_i Y_{t-i} + \sum_{j=1}^{m} \alpha_j X_{t-j} + \varepsilon_t \tag{2.13}$$

Para testar se Y_t Granger-causa X_t, considera-se o seguinte teste:

$$H_0 : \beta_1 = \cdots = \beta_k = 0 \quad \text{vs.} \quad H_1 : \beta_s \neq 0, \text{ para pelo menos um } s \in \{1, \cdots, k\}.$$

Causalidade de Granger e Estacionariedade

Antes de aplicar o método de Causalidade de Granger, precisa-se verificar se as séries são ou não estacionárias, e para isso são utilizados testes como o de Phillips-Perron [Phillips e Perron 1988]. Nesse teste a hipótese nula é de que a série apresenta pelo menos uma raiz unitária (série não estacionária) e a hipótese alternativa é a ausência de raiz unitária. Desse modo, a série será considerada estacionária se rejeitar-se a hipótese nula. Uma análise gráfica preliminar pode auxiliar na decisão da aplicação ou não de testes de estacionariedade. A ausência de tendências determinísticas visíveis e/ou sazonalidades aparentes são índices de estacionariedade. Entretanto, não são suficientes para tomada de decisão, que, preferencialmente, deve ser feita através de um teste apropriado como o Phillips-Perron ou ADF [Dickey e Fuller 1981].

No método de Causalidade de Granger, caso ambas séries forem estacionárias, deve-se seguir as seguintes etapas:

1. Verificar se as séries temporais cointegram utilizando o teste de cointegração de Phillips-Ouliaris [Phillips e Ouliaris 1990].

2. Ajustar um modelo VAR(p), onde p é o número de defasagens. Esse número de defasagens pode ser escolhido utilizando métodos usuais.

3. Aplicar o teste de Causalidade Granger no modelo VAR definido anteriormente. Nele precisa-se declarar qual variável acredita-se que Granger causa a outra. A rejeição da hipótese nula indica a existência de Causalidade de Granger.

Uma forma de aplicar o método de Causalidade de Granger, caso as séries não sejam estacionárias, é utilizar o procedimento de Toda e Yamamoto [Toda e Yamamoto 1995], no qual compreende os seguintes passos:

1. Verificar se as séries cointegram. Duas séries cointegram se possuem a mesma ordem de integração, digamos m, e se o resíduo da regressão de uma série pela outra for estacionário, o que pode ser determinado utilizando-se testes como o de Phillips-Perron.

2. Ajustar um modelo VAR(p).

3. Para aplicar o Teste de Wald, precisa-se ajustar um modelo VAR($p+m$) aos dados. Este modelo incorpora a ordem de integração m às defasagens p, e, assim, permite verificar se os coeficientes são estatisticamente significativos.

Aplicações de Causalidade de Granger

O método de Causalidade de Granger possui ampla aplicabilidade, sendo amplamente explorado, especialmente no campo da economia. Diversos estudos utilizam essa abordagem para investigar relações entre variáveis, oferecendo *insights* relevantes. Alguns exemplos práticos de sua aplicação são apresentados a seguir.

No âmbito econômico, pesquisas analisaram a interdependência entre as principais bolsas de valores globais, avaliando como os mercados interagem entre si e se algum deles exerce influência predominante sobre os outros [Farias e Sáfadi 2010]. Os resultados demonstraram que o mercado brasileiro exerce uma influência significativa sobre os mercados chinês e russo, embora o inverso não tenha sido observado.

Outra aplicação relevante é apresentada por [Dantas e Weydmann 2009], que investigaram a relação entre os preços internos e externos da carne de frango. A pesquisa revelou a existência dessa conexão, evidenciando que os preços internacionais podem ser um indicador útil para o planejamento da produção no mercado interno.

3

Metodologia

Neste capítulo, são descritas as etapas metodológicas seguidas neste trabalho, abrangendo desde a obtenção e preparação dos dados até o desenvolvimento, treinamento e avaliação do modelo LSTM, que constitui a primeira etapa do estudo. Adicionalmente, a segunda etapa aborda a análise das variáveis mais relevantes utilizando métodos de correlação, SHAP e a Causalidade de Granger.

A Figura 3.1 ilustra o fluxograma da primeira etapa do trabalho, que se concentra no desenvolvimento do modelo preditivo LSTM. Este processo inclui desde a coleta e o pré-processamento dos dados até a avaliação do modelo utilizando métricas específicas de desempenho.

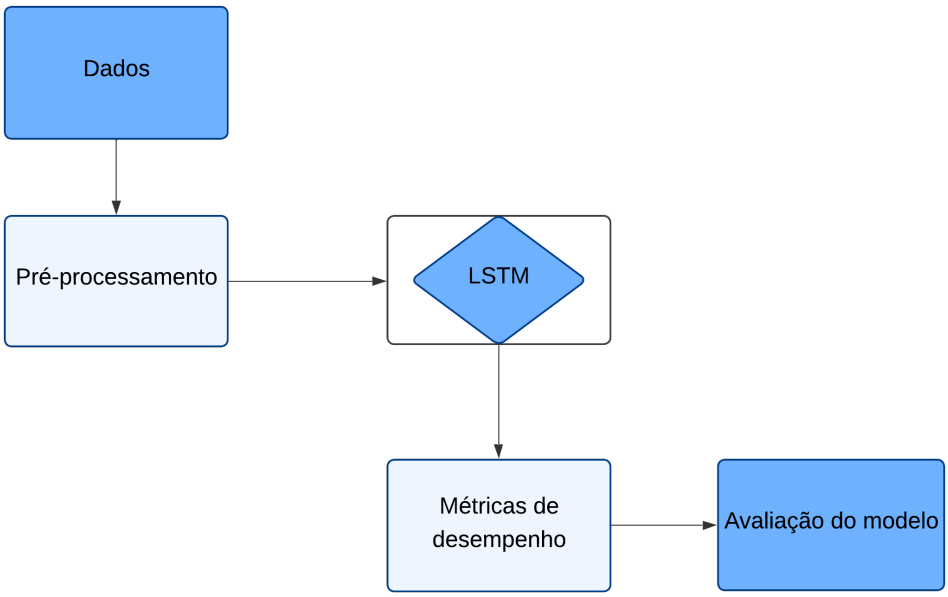

Figura 3.1: Fluxograma da primeira etapa (Autor, 2024).

Posteriormente, a segunda etapa do trabalho, destacada na Figura 3.2, apresenta o fluxo de análise das variáveis de maior impacto. Nessa etapa, foram utilizados métodos de explicação e análise estatística, com o objetivo de identificar relações de Causalidade de Granger e correlação entre as variáveis explicativas e a variável alvo. Essa abordagem permite aprofundar a compreensão dos fatores que influenciam os resultados previstos pelo modelo, além de proporcionar uma interpretação mais transparente e robusta dos dados.

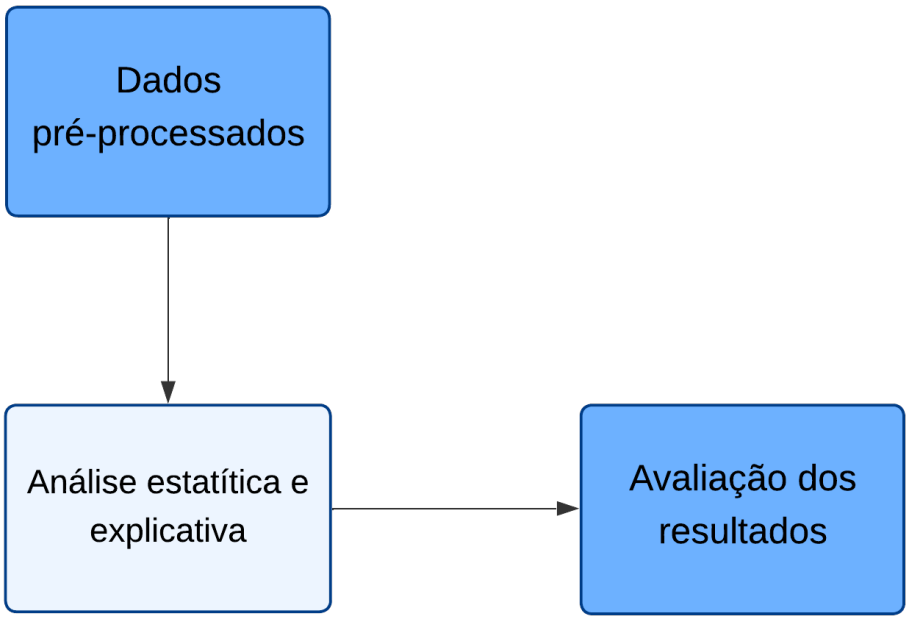

Figura 3.2: Fluxograma da segunda etapa (Autor, 2024).

Com essa estrutura, busca-se proporcionar uma visão detalhada das fases metodológicas, alinhando as etapas práticas do trabalho aos objetivos estabelecidos.

3.1 Base de dados

Os dados utilizados neste estudo são provenientes do **Censo Escolar**, realizado anualmente pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), disponíveis publicamente em seu portal de dados. O conjunto de dados abrange informações detalhadas sobre escolas, turmas, docentes e matrículas de municípios de todo o Brasil. Para este trabalho, a análise foi restrita às **escolas públicas urbanas** localizadas no município de **Maceió, Alagoas**, com dados coletados no período de **2013 a 2022**.

Estrutura e granularidade da base:

- **Granularidade**: Cada linha da base representa uma **escola**, com informações agregadas por ano letivo.
- **Período**: Dados anuais de 2013 a 2022, totalizando **10 anos** de observação.
- **Variáveis principais**: Número de matrículas, infraestrutura escolar, qualificação docente, localização geográfica e indicadores socioeconômicos associados às escolas.
- **Filtros aplicados**:
  - Foram consideradas apenas escolas **públicas** (municipais e estaduais).
  - Excluíram-se unidades inativas ou com dados incompletos no período analisado.
- **Tamanho da amostra**: A base final contém **228 escolas**.

3.2 Pré-processamento dos dados

O pré-processamento dos dados foi realizado com o auxílio da biblioteca scikit-learn, priorizando a normalização das variáveis de entrada. A normalização desempenha um papel fundamental no treinamento de redes neurais, especialmente em arquiteturas como LSTM, que são sensíveis às escalas das variáveis. Esse processo visa garantir que todas as variáveis sejam ajustadas para uma escala comum, eliminando diferenças que possam impactar negativamente o desempenho do modelo.

Os dados foram inicialmente selecionados a partir do Censo Escolar, considerando apenas as variáveis relevantes ao estudo. As variáveis categóricas, foram codificadas numericamente para integração no modelo. Além disso, aplicou-se o método de normalização min-max, que ajusta os valores das variáveis para o intervalo de [0,1]. Essa abordagem é amplamente utilizada em aprendizado de máquina devido à sua capacidade de reduzir a influência de escalas desiguais entre as variáveis.

Uma etapa crucial foi a transformação dos dados em séries temporais, implementada por meio de janelas deslizantes. Esse método organiza os dados em sequências, permitindo que o modelo capture dependências temporais de longo prazo. As janelas deslizantes foram configuradas para dividir as séries temporais em conjuntos de entrada e saída, facilitando a previsão de valores futuros com base em padrões históricos. O resultado desse pré-processamento foi uma base de dados consistente, normalizada e adequada para o treinamento do modelo LSTM.

3.3 Implementação

Os desenvolvimentos computacionais realizados neste trabalho, voltados para o tratamento dos dados e a aplicação de técnicas de *machine learning*, foram implementados utilizando diversas ferramentas e bibliotecas especializadas. A linguagem de programação Python foi o núcleo principal, com destaque para a utilização de bibliotecas como scikit-learn, TensorFlow e Keras, que suportaram os processos de modelagem e treinamento.

O TensorFlow é uma biblioteca Python desenvolvida pela Google que reúne vários modelos e algoritmos de aprendizado de máquina. Sua estrutura é composta por grafos, onde cada nó representa uma operação matemática e cada conexão é uma matriz multidimensional, conhecida como tensor [Yegulalp 2019]. Além disso, ferramentas como SHAP e Statsmodels foram empregadas para interpretação e análise dos modelos, contribuindo para uma compreensão mais aprofundada dos resultados.

Contudo, devido a limitações de compatibilidade do SHAP, foi necessário utilizar versões mais antigas de algumas bibliotecas para garantir seu pleno funcionamento. As bibliotecas e versões utilizadas são apresentadas de forma detalhada na Tabela 3.1.

Ferramentas/Bibliotecas	Scikit-learn (1.3.2)
	Keras/TensorFlow (2.4.1)
	SHAP (0.38.1)
	NumPy (1.19.5)
	Pandas (0.41.0)
	Matplotlib (3.4.3)
	Statsmodels (0.14.1)
Linguagem de Programação	Python (3.8.19)

Tabela 3.1: Ferramentas computacionais utilizadas para o desenvolvimento do trabalho

3.4 Metodologia

3.4.1 Arquitetura da Rede

O modelo utilizado neste trabalho é baseado em redes neurais recorrentes, especificamente na arquitetura *Long Short-Term Memory* (LSTM). Este tipo de rede foi escolhido devido à sua capacidade de capturar dependências de longo prazo em dados temporais.

3.4.2 Construção de Séries Temporais com Janela Deslizante

Para capturar as dependências temporais e permitir que o modelo aprenda padrões históricos relevantes, os dados foram organizados utilizando a técnica de janela deslizante. Nesta aborda-

gem, cada exemplo de entrada é composto por uma sequência de observações consecutivas de um período fixo, e a predição consiste no valor imediatamente subsequente a essa sequência. No presente estudo, optou-se por uma janela deslizante de 3 anos.

Especificamente, para cada ano de previsão, o modelo utiliza os dados referentes aos três anos anteriores. Por exemplo, se tivermos dados de 2013 a 2022, uma janela deslizante de 3 anos gera uma sequência composta pelos dados de 2013, 2014 e 2015 para prever o valor de 2016; em seguida, a janela desliza para englobar 2014, 2015 e 2016 para prever 2017, e assim sucessivamente. Essa abordagem permite que o modelo LSTM incorpore informações históricas suficientes para capturar tendências e variações sazonais ou cíclicas, fundamentais na análise de séries temporais.

A escolha de uma janela de 3 anos baseia-se na hipótese de que os fatores que influenciam o comportamento das matrículas escolares possuem uma persistência temporal que pode ser adequadamente modelada utilizando os dados dos três anos anteriores. Assim, a técnica de janela deslizante não só organiza os dados em um formato compatível com redes LSTM, mas também maximiza a extração de padrões relevantes, contribuindo para a precisão das previsões.

3.4.3 Treinamento

O modelo proposto foi desenvolvido utilizando a biblioteca Keras, amplamente reconhecida por sua flexibilidade e desempenho em tarefas de aprendizado profundo. A arquitetura adotada consistiu em duas camadas LSTM, otimizadas para capturar padrões temporais complexos nos dados.

A primeira camada LSTM foi configurada com 300 unidades e configurada com `return_sequences=True`, permitindo que as saídas fossem encaminhadas para camadas subsequentes. Em seguida, foi aplicada uma camada *Dropout* com taxa de 0,05 para minimizar o risco de *overfitting*. A segunda camada LSTM, composta por 200 unidades e configurada com `return_sequences=False`, foi responsável por sintetizar as informações em uma única saída. Por fim, uma camada densa com uma unidade foi adicionada para realizar a previsão final.

O modelo foi compilado utilizando o otimizador Adam (*Adaptive Moment Estimation*) popularmente utilizado em algoritmos de aprendizado de máquina, especialmente em redes neurais profundas [Kingma e Ba 2014], que combina eficiência computacional e adaptabilidade ao ajustar os pesos da rede, e a função de perda *Mean Squared Error* (MSE), amplamente utilizada para avaliar regressões. O treinamento do modelo foi realizado com um conjunto de dados dividido em 80 para treinamento e 20 para validação. O número máximo de épocas foi fixado em 100, com um tamanho de lote de 256.

Essas configurações garantiram um equilíbrio entre desempenho e eficiência, resultando em um modelo robusto e capaz de capturar as complexidades dos dados analisados.

3.4.4 Teste

A segunda etapa deste trabalho consiste na análise estatística e explicativa das variáveis mais relevantes que influenciam a variável alvo, utilizando métodos como correlação, SHAP (*SHapley Additive Explanations*) e Causalidade de Granger. Esta etapa tem como objetivo principal compreender as relações e dependências entre as variáveis do conjunto de dados, proporcionando uma interpretação mais aprofundada e explicável dos fatores que impactam os resultados do modelo.

Para melhor compreensão dos resultados apresentados neste capítulo, a Tabela 3.2 lista as variáveis utilizadas no estudo, juntamente com seus respectivos significados.

Variável	Descrição
QT_MAT_INF	Quantidade de matrículas no ensino infantil
QT_MAT_FUND	Quantidade de matrículas no ensino fundamental
QT_MAT_MED	Quantidade de matrículas no ensino médio
QT_DOC_INF	Quantidade de docentes no ensino infantil
QT_DOC_FUND	Quantidade de docentes no ensino fundamental
QT_DOC_MED	Quantidade de docentes no ensino médio
QT_TUR_INF	Quantidade de turmas no ensino infantil
QT_TUR_FUND	Quantidade de turmas no ensino fundamental
QT_TUR_MED	Quantidade de turmas no ensino médio
QT_SALAS_EXISTENTES	Quantidade de salas existentes
QT_FUNCIONARIOS	Quantidade de funcionários
QT_COMPUTADOR	Quantidade de computadores
QT_COMP_ALUNO	Quantidade de computadores por aluno
IN_QUADRA_ESPORTES	Existência de quadra esportiva
IN_EQUIP_MULTIMIDIA	Existência de equipamentos multimídia
IN_PARQUE_INFANTIL	Existência de parque infantil
IN_BANHEIRO_EI	Existência de banheiro adequado à educação infantil
IN_EQUIP_RETROPROJETOR	Existência de retroprojetor
IN_BIBLIOTECA_SALA_LEITURA	Existência de biblioteca ou sala de leitura
IN_LABORATORIO_INFORMATICA	Existência de laboratório de informática
IN_LABORATORIO_CIENCIAS	Existência de laboratório de ciências

Tabela 3.2: Variáveis utilizadas no estudo e seus respectivos significados.

Análise de Correlação

A análise de correlação foi realizada para identificar a força e a direção das relações lineares entre as variáveis. Este processo fornece uma visão inicial sobre quais variáveis têm maior

influência nas variáveis-alvo QT_MAT_INF (matrículas no ensino infantil), QT_MAT_FUND (matrículas no ensino fundamental) e QT_MAT_MED (matrículas no ensino médio), permitindo filtrar aquelas que são mais relevantes para análises posteriores.

Os coeficientes de correlação foram calculados utilizando o método de Pearson, que mede a linearidade entre dois conjuntos de dados. Valores próximos de +1 ou -1 indicam uma forte correlação positiva ou negativa, respectivamente, enquanto valores próximos de 0 indicam pouca ou nenhuma correlação.

As variáveis selecionadas para cada nível de ensino, com base na análise de correlação, estão apresentadas na Tabela 3.3.

Nível de Ensino	**Variável-Alvo**	**Variáveis Relevantes**	**Coeficiente de Correlação**
Ensino Infantil	QT_MAT_INF	QT_DOC_INF	0.88
		QT_TUR_INF	0.85
		IN_EQUIP_MULTIMIDIA	0.56
		IN_BANHEIRO_EI	0.64
		IN_PARQUE_INFANTIL	0.56
		IN_LABORATORIO_INFORMATICA	0.55
Ensino Fundamental	QT_MAT_FUND	QT_DOC_FUND	0.89
		QT_TUR_FUND	0.82
		QT_SALAS_EXISTENTES	0.60
		QT_FUNCIONARIOS	0.65
		IN_BIBLIOTECA_SALA_LEITURA	0.56
		IN_EQUIP_RETROPROJETOR	0.55
Ensino Médio	QT_MAT_MED	QT_DOC_MED	0.85
		QT_TUR_MED	0.83
		IN_LABORATORIO_CIENCIAS	0.58
		QT_FUNCIONARIOS	0.56
		QT_COMPUTADOR	0.60
		QT_COMP_ALUNO	0.58
		IN_QUADRA_ESPORTES	0.55

Tabela 3.3: Análise das variáveis relevantes e suas correlações

De forma geral, observa-se uma forte correlação positiva entre as variáveis relevantes e a variável alvo (número de matrículas) em todos os níveis de ensino (Infantil, Fundamental e Médio). Isso indica que as variáveis selecionadas desempenham um papel importante na previsão do número de matrículas.

Interpretação com SHAP

Após identificar as variáveis relevantes por meio da correlação, foi aplicado o método SHAP para interpretar o impacto de cada variável nos resultados do modelo. O SHAP, baseado na teoria dos valores de Shapley, fornece uma medida quantitativa da contribuição de cada variável

para a predição final do modelo.

- *Beeswarm Plot*: Este gráfico foi utilizado para visualizar a densidade das contribuições de cada variável. Ele permite identificar quais variáveis possuem maior impacto nas predições ao longo do conjunto de dados.
- Importância Global: A soma das magnitudes dos valores SHAP foi utilizada para ranquear as variáveis em ordem de importância, destacando aquelas com maior relevância no modelo.

Análise de Causalidade com Granger

Para investigar as relações de causa e efeito entre as variáveis, foi aplicada a Causalidade de Granger. Este método estatístico avalia se uma variável pode ser utilizada para prever outra, considerando séries temporais.

Foram avaliadas todas as combinações de variáveis dependentes e independentes para determinar relações significativas de Causalidade de Granger. As variáveis com valores-p abaixo do nível de significância (geralmente 0,05) foram consideradas como causalmente relacionadas.

3.4.5 Análise Explicativa e de Causalidade de Granger

A apresentação dos resultados deste estudo iniciou-se com a organização e análise dos dados provenientes das etapas de treinamento e validação do modelo LSTM. Esses resultados refletiram o desempenho do modelo ao prever as matrículas em séries temporais, considerando as variáveis selecionadas. A estruturação dos dados e das previsões possibilitou uma análise detalhada e abrangente, permitindo identificar padrões e tendências significativas no comportamento dos dados ao longo do tempo.

Os resultados foram avaliados com base em métricas amplamente aceitas na área de aprendizado de máquina para problemas de regressão, como o erro quadrático médio (MSE), erro absoluto médio (MAE) e o coeficiente de determinação (R^2). Essas métricas permitiram medir a precisão e a eficácia do modelo, avaliando o alinhamento entre as previsões e os valores reais observados. Além disso, análises complementares foram realizadas utilizando o SHAP para interpretar a importância das variáveis no modelo, fornecendo uma visão explicativa dos fatores que mais influenciaram as matrículas previstas.

Adicionalmente, a segunda etapa do trabalho abordou a análise de correlação e Causalidade de Granger. O método de Causalidade de Granger foi aplicado para verificar relações de dependência entre as variáveis consideradas, enquanto o SHAP forneceu *insights* sobre a contribuição de cada variável para os resultados do modelo. Essa combinação de métodos estatísticos e explicativos garantiu uma análise robusta e fundamentada dos fatores determinantes para as previsões realizadas.

Por fim, a apresentação dos resultados contemplou uma análise comparativa das variáveis relacionadas às matrículas em diferentes etapas de ensino (infantil, fundamental e médio), destacando as diferenças de impacto entre elas. Essa abordagem permitiu uma compreensão mais profunda do comportamento das variáveis ao longo do tempo e em diferentes contextos educacionais, enriquecendo as discussões e contribuindo para o entendimento das dinâmicas envolvidas no problema.

4

Resultados

Neste capítulo, são apresentados os resultados obtidos a partir da análise das séries temporais referentes às matrículas escolares no município de Maceió da rede pública. As análises foram realizadas considerando os diferentes níveis de ensino — Infantil, Fundamental e Médio — com o objetivo de identificar os fatores que mais influenciam o comportamento das matrículas ao longo do tempo.

Inicialmente, é realizada uma decomposição das séries temporais para explorar seus componentes de tendência, sazonalidade e ruído. Em seguida, os modelos de previsão baseados em LSTM são aplicados para cada nível de ensino, e as métricas de desempenho como MAE, MSE, RMSE e R^2 são calculadas para avaliar a precisão das previsões.

Por fim, as seções dedicadas a cada nível de ensino detalham as principais descobertas, correlacionando os resultados obtidos com a literatura existente e destacando as implicações para políticas educacionais e gestão escolar.

4.1 Ensino Infantil

A Figura 4.1 ilustra o desempenho do modelo na previsão de matrículas para o Ensino Infantil. Observa-se um razoável alinhamento entre as linhas de valores previstos e reais, indicando que o modelo consegue capturar parte da tendência das matrículas neste segmento. No entanto, existem alguns pontos de divergência, sugerindo limitações do modelo em prever flutuações mais abruptas nos dados.

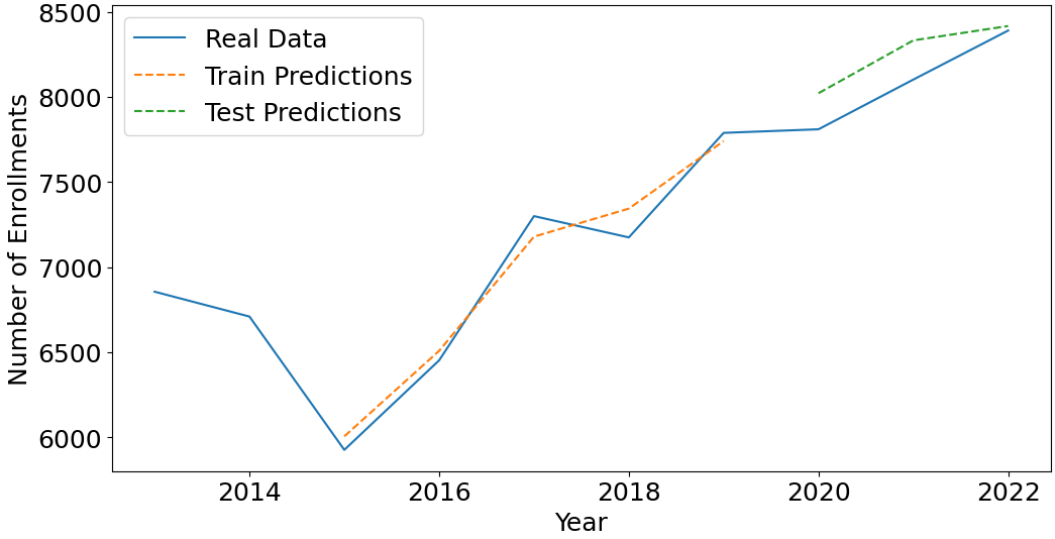

Figura 4.1: Comparação entre matrículas reais e previstas - Ensino Infantil (Autor, 2024).

A Tabela 4.1 complementa a análise visual, fornecendo as métricas de desempenho do modelo para o Ensino Infantil.

Métrica	**Valor**
MAE	137.41
RMSE	166.32
MSE	27663.03
R^2	0.51

Tabela 4.1: Métricas de desempenho do modelo - Ensino Infatil (Autor, 2024).

- **MAE**: 137.41 (Erro Médio Absoluto). O modelo erra em média 137 matrículas, representando aproximadamente 2% de erro relativo (média de 7000 matrículas), resultando em uma média ligeiramente superior ou inferior a 2%.
- **RMSE**: O RMSE de 166.32 matrículas indica que, em média, as previsões do modelo desviam-se dos valores reais em 166.32 matrículas. Esta métrica é sensível a erros maiores, penalizando-os mais fortemente.
- **MSE**: O MSE de 27663.03 representa a média dos erros quadrados das previsões. É uma métrica útil para comparar o desempenho de diferentes modelos e para otimização.
- $\mathbf{R}^2$: 0.51 (Coeficiente de Determinação). O modelo explica 51% da variabilidade dos dados.

As Figuras 4.2 e 4.3 apresentam análises de interpretabilidade no qual foi conduzida usando valores SHAP para identificar as variáveis que mais influenciam o número de matrículas do ensino infantil.

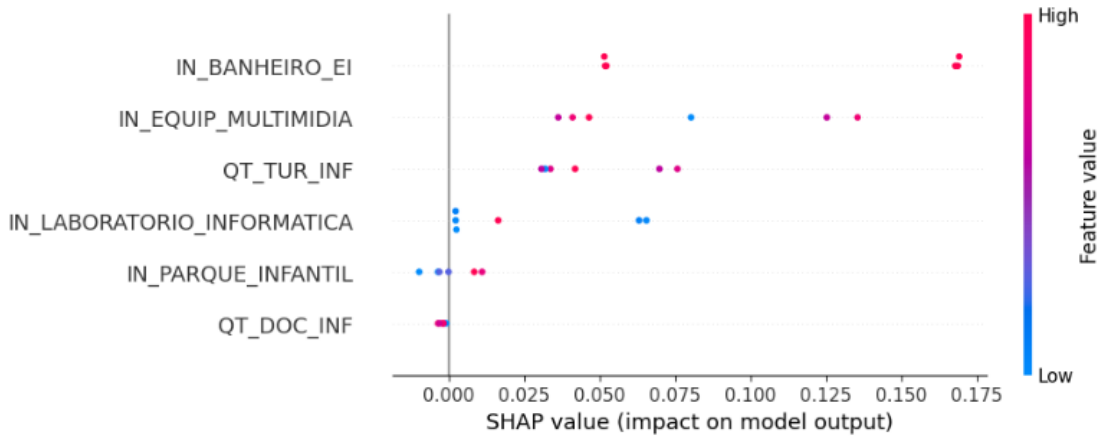

Figura 4.2: *Beeswarm Plot* - Ensino Infantil (Autor, 2024).

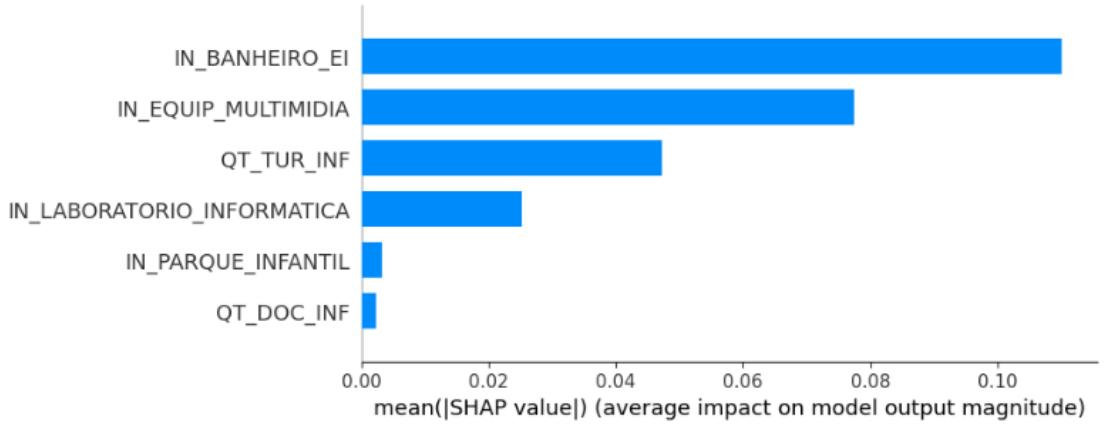

Figura 4.3: *Feature Importance* - Ensino Infantil (Autor, 2024).

A análise revela que `IN_BANHEIRO_EI` (existência de banheiro adequado para a Educação Infantil) é a variável mais influente. Escolas com essa infraestrutura (pontos vermelhos) tendem a ter valores SHAP positivos (à direita), indicando um impacto positivo nas matrículas, enquanto escolas sem essa estrutura (pontos azuis) apresentam valores SHAP negativos (à esquerda) ou próximos de zero, sugerindo um impacto negativo ou nulo. A variável `IN_EQUIP_MULTIMIDIA` (presença de equipamentos multimídia) aparece como a segunda variável mais importante, com um padrão semelhante: escolas com mais equipamentos multimídia (pontos vermelhos) correlacionam-se com valores SHAP positivos, sugerindo um impacto positivo na atratividade da escola. A ausência desses equipamentos (pontos azuis) está associada a valores SHAP mais baixos.

A variável `QT_TUR_INF` (quantidade de turmas no ensino infantil) surge como a terceira variável em importância. Um maior número de turmas (pontos vermelhos) está associado a valores SHAP positivos, indicando que escolas com maior oferta de turmas tendem a apresentar mais matrículas. A variável `IN_LABORATORIO_INFORMATICA` (presença de laboratório de informática) aparece na sequência, com uma influência moderada. A presença de laboratórios de

informática está associada a valores SHAP positivos, embora haja uma dispersão nos valores, sugerindo que sua contribuição pode variar conforme o contexto da escola.

A variável IN_PARQUE_INFANTIL (presença de parque infantil) apresenta uma relação menos expressiva, mas ainda positiva. A presença dessa estrutura pode contribuir para o aumento das matrículas, embora com menor peso comparado às demais variáveis. A variável QT_DOC_INF (quantidade de docentes no ensino infantil) aparece com a menor importância no modelo. Diferentemente do ensino fundamental, onde o número de docentes tem um impacto mais relevante, na Educação Infantil essa variável não se mostra decisiva na previsão das matrículas.

A Figura 4.4 apresenta os resultados da análise de Causalidade de Granger em uma representação de grafo bipartido, no qual os nós em azul representam as variáveis independentes e os nós em laranja representam as variáveis dependentes. Isso permite verificar se as variáveis identificadas como relevantes pelo SHAP também influenciam outras variáveis ao longo do tempo.

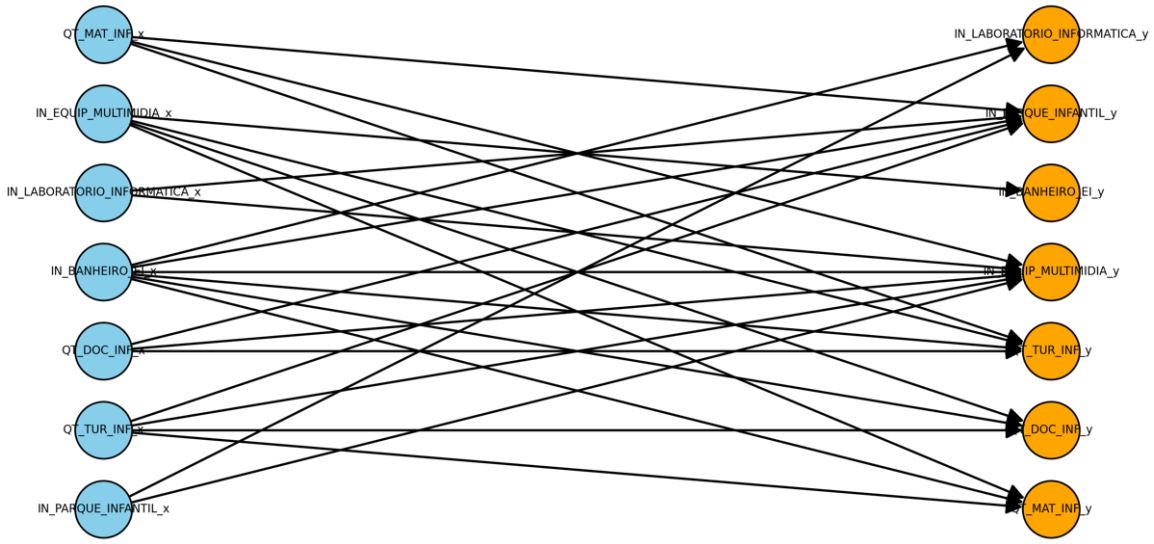

Figura 4.4: Causalidade de Granger - Ensino Infantil (Autor, 2024).

Para QT_MAT_INF_x, há evidência de Causalidade de Granger para as variáveis QT_TUR_INF_y (p=0,0006), IN_EQUIP_MULTIMIDIA_y (p=0,0000) e IN_PARQUE_INFANTIL_y (p=0,0000). Não se verifica relação estatisticamente significativa com banheiros infantis, laboratórios de informática e docentes.

No caso de QT_DOC_INF_x, p-valores significativos em IN_EQUIP_MULTIMIDIA_y (p=0,0000) QT_TUR_INF_y (p=0,0000) e IN_PARQUE_INFANTIL_y (p=0,0000). Em contrapartida, não há evidência de Causalidade de Granger em matrículas, banheiros e laboratórios (p=0,4776, 0,4187 e 0,8465, respectivamente).

Para QT_TUR_INF_x, destaca-se efeito sobre QT_MAT_INF_y (p=0,0000), QT_DOC_INF_y (p=0,0000), IN_EQUIP_MULTIMIDIA_y (p=0,0000) e IN_PARQUE_INFANTIL_y (p=0,0019).

Não se observa Causalidade de Granger em banheiros infantis (p=0,2302) nem em laboratórios (p=0,0525).

Para IN_EQUIP_MULTIMIDIA_x, observam-se p-valores muito baixos ao explicar QT_MAT_INF_y (p=0,0000), QT_DOC_INF_y (p=0,0000) e QT_TUR_INF_y (p=0,0000), indicando que a disponibilidade de equipamentos multimídia pode influenciar matrículas, docentes e turmas. Também há significância estatística em relação a IN_BANHEIRO_EI_y (p=0,0378), revelando possível impacto de equipamentos multimídia sobre a presença de banheiros na Educação Infantil. Em contrapartida, não se detecta Causalidade de Granger para IN_PARQUE_INFANTIL_y (p=0,4269) nem IN_LABORATORIO_INFORMATICA_y (p=0,2156).

No caso de IN_BANHEIRO_EI_x, os p-valores também são significativos em QT_MAT_INF_y (p=0,0000), QT_DOC_INF_y (p=0,0000) e QT_TUR_INF_y (p=0,0000), apontando influência sobre matrículas, docentes e turmas. Destacam-se ainda IN_EQUIP_MULTIMIDIA_y (p=0,0000), IN_PARQUE_INFANTIL_y (p=0,0098) e IN_LABORATORIO_INFORMATICA_y (p=0,0011), o que sugere influência dos banheiros sobre diversos recursos escolares.

Para IN_PARQUE_INFANTIL_x, os p-valores também são significativos em IN_LABORATORIO_INFORMATICA_y (p=0,0455) e IN_EQUIP_MULTIMIDIA_y (p=0,0155), nas demais variáveis não se observa Causalidade de Granger.

Por fim, IN_LABORATORIO_INFORMATICA_x, os p-valores também são significativos em IN_PARQUE_INFANTIL_y (p=0,0157) e IN_EQUIP_MULTIMIDIA_y (p=0,0009). Nos demais cruzamentos, como matrículas, docentes, turmas ou a existência de banheiro infatil, os p-valores não indicam associação estatisticamente significativa.

4.2 Ensino Fundamental

A Figura 4.5 ilustra o desempenho do modelo na previsão de matrículas para o Ensino Fundamental. Observa-se um alinhamento considerável entre as linhas de valores previstos e reais, especialmente nos anos iniciais e finais da série histórica. Isso indica uma boa capacidade do modelo em capturar a tendência geral das matrículas neste segmento.

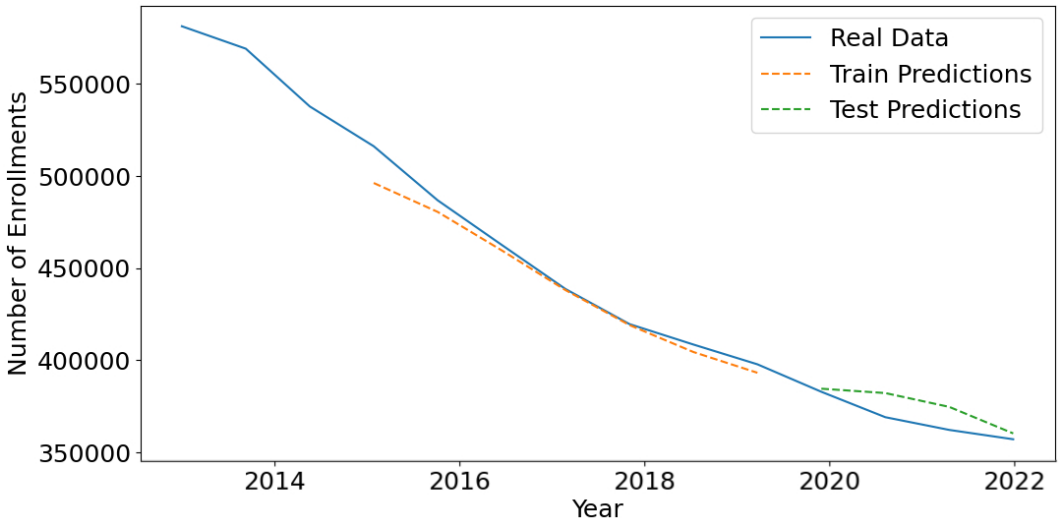

Figura 4.5: Comparação entre matrículas reais e previstas - Ensino Fundamental (Autor, 2024).

A Tabela 4.2 fornece as métricas de desempenho do modelo ao contexto do Ensino Fundamental e complementa a análise visual.

Métrica	**Valor**
MAE	6534.19
RMSE	7708.74
MSE	59424611.42
R^2	0.72

Tabela 4.2: Métricas de desempenho do modelo - Ensino Fundamental (Autor, 2024)

- **MAE**: 6534.19 (Erro Médio Absoluto). Erro médio de 1.6% (média de 400.000 matrículas). Assim como no caso anterior, a média dos erros absolutos pode variar ligeiramente, situando o erro relativo um pouco acima ou abaixo de 1.6%.
- **RMSE**: O RMSE de 7708.74 matrículas indica que, em média, as previsões do modelo desviam-se dos valores reais em 7708.74 matrículas. Esta métrica penaliza erros maiores.
- **MSE**: O MSE de 59424611.42 representa a média dos erros quadrados das previsões, sendo útil para comparação de modelos.
- **R^2**: 0.72 (Coeficiente de Determinação). Explica 72% da variação dos dados de matrículas.

As Figuras 4.6 e 4.7 apresentam as análises de interpretabilidade para o Ensino Fundamental, utilizando valores SHAP para avaliar o impacto de diferentes variáveis no número de matrículas.

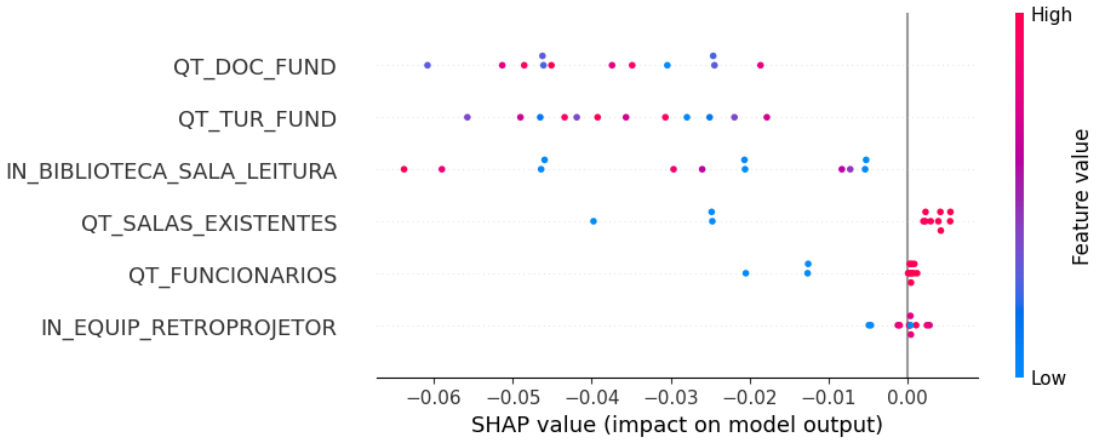

Figura 4.6: *Beeswarm Plot* - Ensino Fundamental (Autor, 2024).

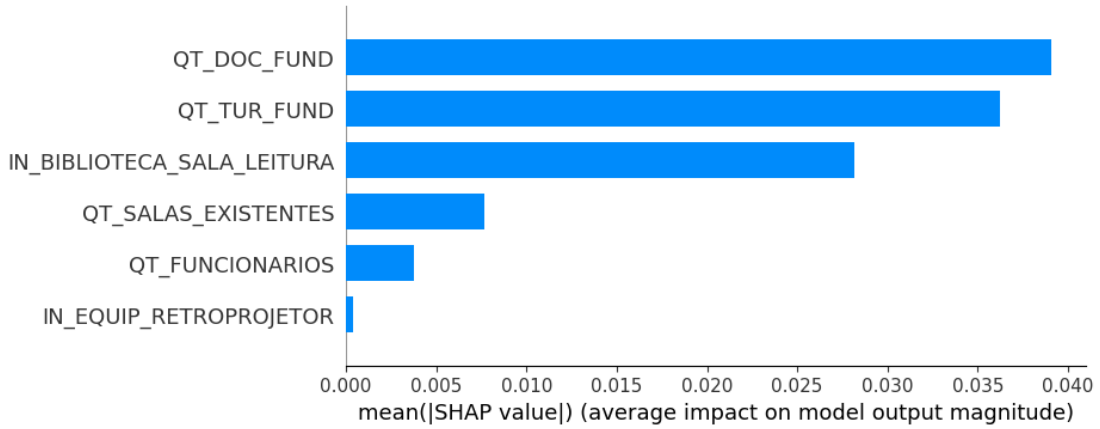

Figura 4.7: *Feature Importance* - Ensino Fundamental (Autor, 2024).

A análise revela que `QT_DOC_FUND` (quantidade de docentes do ensino fundamental) é a variável mais influente. Escolas com mais docentes (pontos vermelhos) tendem a ter valores SHAP positivos (à direita), indicando um aumento nas matrículas, enquanto escolas com menos docentes (pontos azuis) apresentam valores SHAP negativos (à esquerda) ou próximos de zero, sugerindo um impacto negativo ou nulo. A variável `QT_TUR_FUND` (quantidade de turmas do ensino fundamental) aparece como a segunda variável mais importante, com um padrão semelhante: mais turmas (pontos vermelhos) correlacionam-se com valores SHAP positivos, impactando positivamente as matrículas.

A variável `IN_BIBLIOTECA_SALA_LEITURA` (existência de biblioteca ou sala de leitura) surge como a terceira variável em importância. A presença de biblioteca ou sala de leitura (pontos vermelhos) está associada a valores SHAP positivos, embora haja alguma dispersão, sugerindo uma contribuição positiva para a atratividade da escola.

A variável `QT_SALAS_EXISTENTES` (quantidade de salas de aula existentes) vem na sequência, com uma influência mais moderada, mas ainda positiva: mais salas (pontos vermelhos) tendem a valores SHAP ligeiramente positivos.

A variável QT_FUNCIONARIOS (quantidade de funcionários) apresenta uma relação menos clara, com pontos vermelhos e azuis distribuídos próximos ao zero, indicando um impacto baixo ou nulo no número de matrículas. Por fim, a variável IN_EQUIP_RETROPROJETOR (existência de retroprojetor) demonstra ser a menos influente, com a maioria dos pontos agrupada em torno de zero, indicando impacto mínimo ou inexistente.

A Figura 4.8 apresenta os resultados da análise de Causalidade de Granger para o Ensino Fundamental em uma representação de grafo bipartido, no qual os nós em azul representam as variáveis independentes e os nós em laranja representam as variáveis dependente, permitindo identificar relações temporais entre as variáveis analisadas.

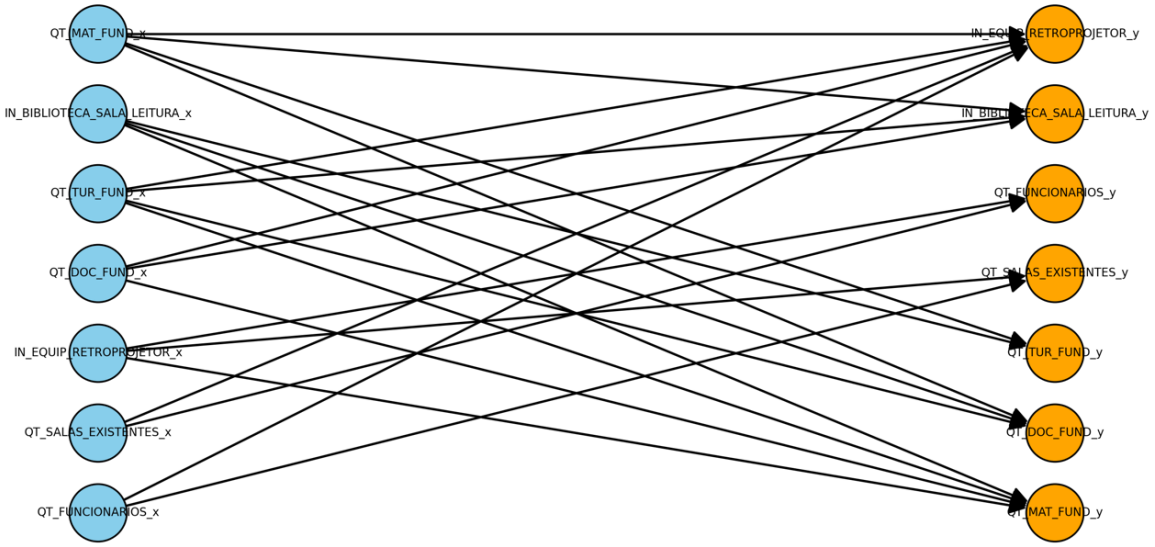

Figura 4.8: Causalidade de Granger - Ensino Fundamental (Autor, 2024).

Para IN_BIBLIOTECA_SALA_LEITURA_x, observam-se p-valores abaixo de 0,05 em QT_MAT_FUND_y (p=0,0062), QT_DOC_FUND_y (p=0,0129) e QT_TUR_FUND_y (p=0,0001), indicando que a presença histórica de bibliotecas ou salas de leitura Granger-causa o aumento de matrículas, docentes e turmas. Em contrapartida, não se verifica Causalidade de Granger estatisticamente significativa em QT_SALAS_EXISTENTES_y (p=0,5143), QT_FUNCIONARIOS_y (p=0,4274), IN_EQUIP_RETROPROJETOR_y (p=0,1460).

Para IN_EQUIP_RETROPROJETOR_x, os p-valores confirmam efeito significativo sobre QT_MAT_FUND_y (p=0,0466), QT_SALAS_EXISTENTES_y (p=0,0001) e QT_FUNCIONARIOS_y (p=0,0000), sugerindo que o uso de retroprojetores influi em matrículas, número de salas e número de funcionários. Entretanto, não há evidências de Causalidade de Granger em relação a QT_DOC_FUND_y (p=0,3504), QT_TUR_FUND_y (p=0,0907), IN_BIBLIOTECA_SALA_LEITURA_y (p=0,3452).

Para QT_MAT_FUND_x, observam-se valores de p significativos ao explicar QT_DOC_FUND_y (p=0,0000), QT_TUR_FUND_y (p=0,0000), IN_BIBLIOTECA_SALA_LEITURA_y (p=0,0170) e IN_EQUIP_RETROPROJETOR_y (p=0,0175), o que indica relação de Granger-causalidade das

matrículas sobre docentes, turmas e certos recursos de infraestrutura. Não há evidência de Causalidade de Granger para QT_SALAS_EXISTENTES_y (p=0,1147) nem QT_FUNCIONARIOS_y (p=0,0851).

Já QT_DOC_FUND_x se mostra relevante para QT_MAT_FUND_y (p=0,0266), IN_BIBLIOTECA_SALA_LEITURA_y (p=0,0164) e IN_EQUIP_RETROPROJETOR_y (p=0,0112), mas não se relaciona a QT_TUR_FUND_y (p=0,0710), QT_SALAS_EXISTENTES_y (p=0,0918) nem com a quantidade de funcionários QT_FUNCIONARIOS_y (p=0,0656).

Para QT_TUR_FUND_x, os valores de p confirmam que turmas Granger-causam QT_MAT_FUND_y (p=0,0006) e QT_DOC_FUND_y (p=0,0001), além de IN_BIBLIOTECA_SALA_LEITURA_y (p=0,0000) e IN_EQUIP_RETROPROJETOR_y (p=0,0087). Por outro lado, não há Causalidade de Granger para QT_SALAS_EXISTENTES_y (p=0,0744) e QT_FUNCIONARIOS_y (p=0,0527), que exibem p-valores acima do limiar de significância.

Para QT_SALAS_EXISTENTES_x, observa-se relação de Granger-causalidade com as variáveis QT_FUNCIONARIOS_y (p=0,0298) e IN_EQUIP_RETROPROJETOR_y (p=0,0001), indicando que a quantidade de salas existentes pode influenciar tanto o número de funcionários quanto a adoção de retroprojetores. Em contrapartida, não há evidência de Causalidade de Granger para QT_MAT_FUND_y, QT_DOC_FUND_y, QT_TUR_FUND_y e IN_BIBLIOTECA_SALA_LEITURA_y, cujos p-valores excedem o limite de significância.

No caso de QT_FUNCIONARIOS_x, verifica-se Causalidade de Granger em relação a QT_SALAS_EXISTENTES_y (p=0,0294) e IN_EQUIP_RETROPROJETOR_y (p=0,0000), ou seja, o histórico de funcionários parece influenciar o número de salas disponíveis e a presença de retroprojetores. Nos demais cruzamentos, como matrículas, docentes, turmas ou biblioteca/sala de leitura, os p-valores não indicam associação estatisticamente significativa.

4.3 Ensino Médio

A Figura 4.9 ilustra o desempenho do modelo na previsão de matrículas para o Ensino Médio. Nota-se um alinhamento expressivo entre as linhas de valores previstos e reais, revelando um bom ajuste do modelo aos dados e sua capacidade de capturar as tendências das matrículas neste segmento educacional.

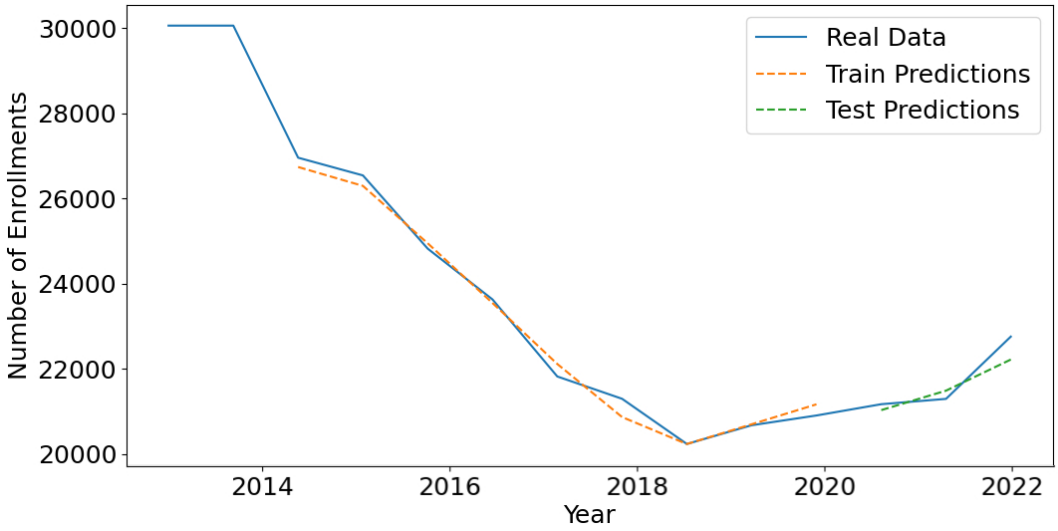

Figura 4.9: Comparação entre matrículas reais e previstas - Ensino Médio (Autor, 2024).

A Tabela 4.3 fornece as métricas de desempenho do modelo aplicado especificamente aos dados do Ensino Médio.

Métrica	**Valor**
MAE	247.43
RMSE	271.24
MSE	73569.59
R^2	0.86

Tabela 4.3: Métricas de desempenho do modelo - Ensino Médio (Autor, 2024)

- **MAE**: 247.43 (Erro Médio Absoluto). O modelo erra em média 247.43 matrículas (1%), assim como nos casos anteriores, a média dos erros absolutos pode variar ligeiramente, situando o erro relativo um pouco acima ou abaixo de 1%.
- **RMSE**: 271.24 (Raiz do Erro Quadrático Médio). As previsões do modelo desviam-se dos valores reais em média 271.24 matrículas, com penalização para erros maiores.
- **MSE**: 73569.59 (Erro Quadrático Médio). A média dos erros quadrados das previsões é 73569.59, útil para comparação de modelos.
- **R^2**: 0.86 (Coeficiente de Determinação). O modelo explica 86% da variabilidade dos dados.

As Figuras 4.10 e 4.11 apresentam as análises de interpretabilidade para o Ensino Médio, utilizando valores SHAP para avaliar o impacto de diferentes variáveis no número de matrículas.

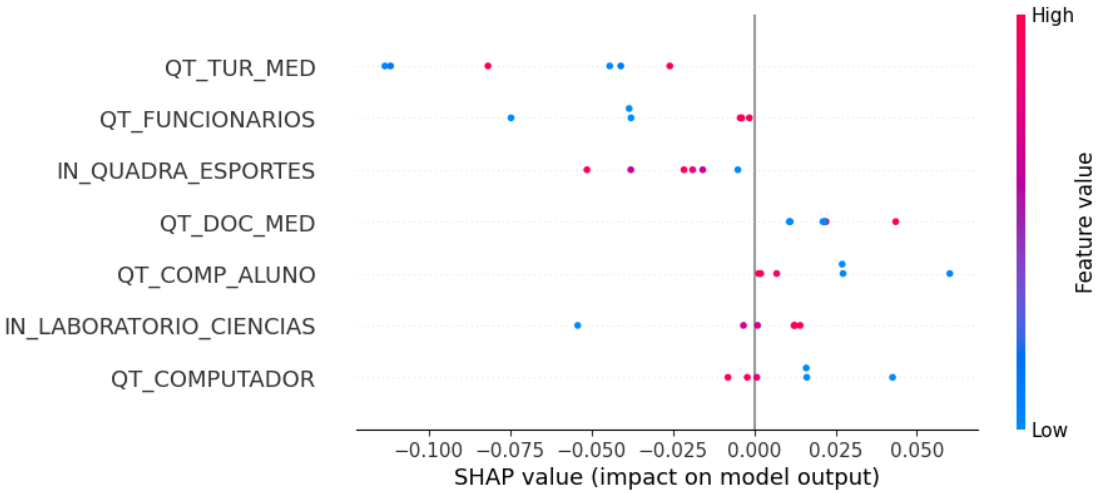

Figura 4.10: *Beeswarm Plot* - Ensino Médio (Autor, 2024).

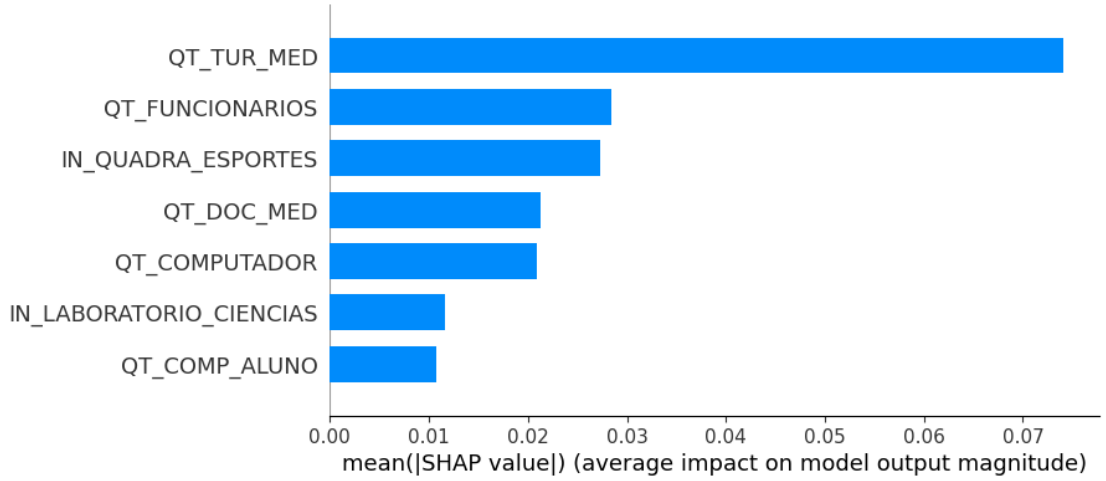

Figura 4.11: *Feature Importance* - Ensino Médio (Autor, 2024).

A análise revela que `QT_TUR_MED` (quantidade de turmas do ensino médio) é a variável mais influente. Escolas com mais turmas (pontos vermelhos) tendem a ter valores SHAP positivos (à direita), indicando um aumento nas matrículas, enquanto escolas com menos turmas (pontos azuis) apresentam valores SHAP negativos (à esquerda) ou próximos de zero, sugerindo um impacto negativo ou nulo. A variável `QT_FUNCIONARIOS` (quantidade de funcionários) aparece como a segunda mais importante, com um padrão semelhante: mais funcionários (pontos vermelhos) correlacionam-se com valores SHAP positivos, impactando positivamente as matrículas.

A variável `IN_QUADRA_ESPORTES` (existência de quadra de esportes) surge como a terceira em importância. Embora a dispersão seja menor que nas variáveis anteriores, a presença de quadra (pontos vermelhos) está mais associada a valores SHAP positivos, sugerindo uma contribuição positiva para a atratividade da escola. Já a variável `QT_DOC_MED` (quantidade de do-

centes do ensino médio) apresenta uma influência mais moderada, mas ainda positiva: mais professores (pontos vermelhos) tendem a valores SHAP ligeiramente positivos.

A variável QT_COMPUTADOR (quantidade de computadores) apresenta uma relação similar a variável anterior, porém com menor impacto no número de matrículas. A variável IN_LABORATORIO_CIENCIAS (existência de laboratório de ciências) mostra uma influência sutil: a presença de laboratório (pontos vermelhos) tende a valores SHAP ligeiramente positivos, mas a dispersão é pequena. Por fim, a variável QT_COMP_ALUNO (quantidade de computadores por aluno) demonstra ser a menos influente, com a maioria dos pontos agrupada em torno de zero, indicando impacto mínimo ou inexistente.

A Figura 4.12 apresenta os resultados da análise de Causalidade de Granger para o Ensino Médio em uma representação de grafo bipartido, no qual os nós em azul representam as variáveis independentes e os nós em laranja representam as variáveis dependente, permitindo identificar relações temporais entre as variáveis analisadas.

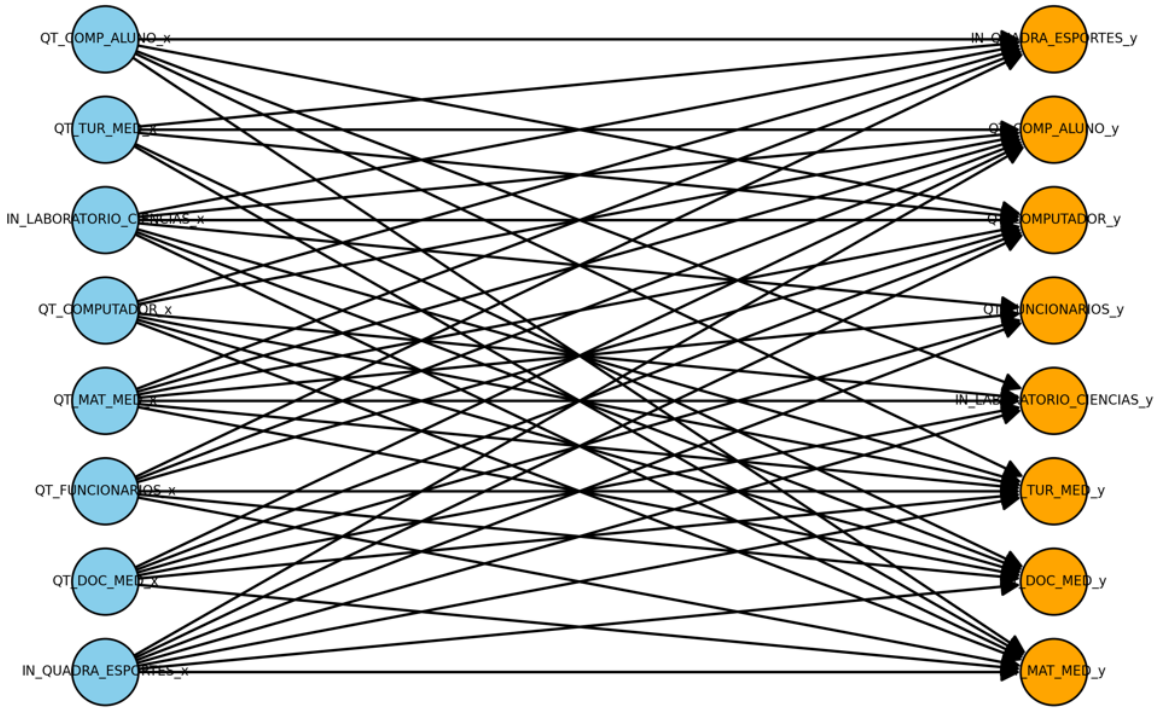

Figura 4.12: Causalidade de Granger - Ensino Médio (Autor, 2024).

Para QT_MAT_MED_x, verificam-se p-valores menores que 0,05 em praticamente todas as combinações, exceto na própria QT_MAT_MED_y (p=1,0000). Assim, há evidência de Granger-causalidade sobre QT_DOC_MED_y (p=0,0004), QT_TUR_MED_y (p=0,0040), IN_LABORATORIO_CIENCIAS_y (p=0,0309), QT_FUNCIONARIOS_y (p=0,0008), QT_COMPUTADOR_y (p=0,0011), QT_COMP_ALUNO_y (p=0,0000) e IN_QUADRA_ESPORTES_y (p=0,0001). Esses resultados indicam que o histórico de matrículas exerce influência importante sobre docentes, turmas, funcionários, computadores e outros recursos, bem como na presença de laboratórios de ciências e

quadras esportivas.

Para QT_DOC_MED_x, nota-se Granger-causalidade nas variáveis QT_MAT_MED_y (p=0,0000), QT_TUR_MED_y (p=0,0000), IN_LABORATORIO_CIENCIAS_y (p=0,0001), QT_FUNCIONARIOS_y (p=0,0171), QT_COMPUTADOR_y (p=0,0000) e QT_COMP_ALUNO_y (p=0,0000). Entretanto, não há relação estatisticamente significativa com IN_QUADRA_ESPORTES_y (p=0,0748). Dessa forma, a quantidade de docentes no Ensino Médio aparenta influenciar matrículas, turmas, laboratório de ciências e equipamentos de TI, além do corpo de funcionários.

Quanto a IN_LABORATORIO_CIENCIAS_x, evidencia-se p-valores significativos paras todas as demais variáveis QT_MAT_MED_y (p=0,0000), QT_DOC_MED_y (p=0,0019), QT_TUR_MED_y (p=0,0001), QT_COMP_ALUNO_y (p=0,0126), QT_COMPUTADOR_y (p=0,0020), QT_FUNCIONARIOS_y (p=0,0000) e IN_QUADRA_ESPORTES_y (p=0,0000). Isto indica que a presença de laboratório de ciências possui forte influencia nos últimos anos da escola.

Quanto a QT_FUNCIONARIOS_x, evidencia-se p-valores significativos paras as variáveis QT_MAT_MED_y (p=0,0000), QT_DOC_MED_y (p=0,0000), QT_TUR_MED_y (p=0,0012), IN_QUADRA_ESPORTES_y (p=0,0005). QT_COMP_ALUNO_y (p=0,0013) e QT_COMPUTADOR_y (p=0,0000). Por outro lado, não há indícios Causalidade de Granger para a variável IN_LABORATORIO_CIENCIAS_y (p=0,2738).

Quanto a QT_TUR_MED_x, evidencia-se p-valores significativos para a variável QT_MAT_MED_y (p=0,0001), QT_DOC_MED_y (p=0,0352), QT_COMPUTADOR_y (p=0,0033), QT_COMP_ALUNO_y (p=0,0029) e IN_QUADRA_ESPORTES_y (p=0,0029). Por outro lado, não há Causalidade de Granger para a variável IN_LABORATORIO_CIENCIAS_y (p=0,1532) nem QT_FUNCIONARIOS_y (p=0,2761). Isso sugere que o número de turmas influencia variáveis como matrículas, docentes, computadores, razão computador/aluno e quadra esportiva, mas não apresenta relação estatisticamente significativa com o laboratório de ciências ou com a quantidade de funcionários no período analisado.

Para QT_COMPUTADOR_x, observa-se relação de Granger-causalidade com a variável QT_MAT_MED_y (p=0,0000), QT_DOC_MED_y (p=0,0152), QT_TUR_MED_y (p=0,0000), IN_LABORATORIO_CIENCIAS_y (p=0,0000), QT_COMP_ALUNO_y (p=0,0031) e para IN_QUADRA_ESPORTES_y (p=0,0007). Por outro lado, não há indícios de Causalidade de Granger para a variável QT_FUNCIONARIOS_y (p=0,0664).

Para QT_COMP_ALUNO_x, p-valores significativos em QT_MAT_MED_y (p=0,0000), QT_TUR_MED_y (p=0,0000), IN_LABORATORIO_CIENCIAS_y (p=0,0000), QT_COMPUTADOR_y (p=0,0033) e IN_QUADRA_ESPORTES_y (p=0,0022) apontam influência sobre essas variáveis. Já QT_DOC_MED_y (p=0,0766), QT_FUNCIONARIOS_y (p=0,2598).

Quanto a IN_QUADRA_ESPORTES_x, os p-valores mostram Causalidade de Granger para as variáveis QT_MAT_MED_y (p=0,0000), QT_DOC_MED_y (p=0,0000), QT_TUR_MED_y (p=0,0000), IN_LABORATORIO_CIENCIAS_y (p=0,0258), QT_FUNCIONARIOS_y (p=0,0011), QT_COMPUTADOR_y (p=0,0004) e QT_COMP_ALUNO_y (p=0,0009).

4.4 Discussão dos Resultados

Os resultados indicam que a quantidade de matrículas nos ensinos Infantil, Fundamental e Médio não depende apenas de fatores isolados, mas de um conjunto de interações entre infraestrutura, corpo docente e disponibilidade de recursos pedagógicos e tecnológicos. Além disso, algumas variáveis apresentam relações entre si, o que reforça a necessidade de uma gestão escolar integrada, onde decisões sobre investimentos e cortes orçamentários devem considerar impactos diretos e indiretos no funcionamento da escola.

No **ensino infantil**, a infraestrutura da escola se mostrou essencial não apenas para a atratividade da instituição, mas também para a manutenção de um ambiente adequado para o desenvolvimento das crianças. A presença de banheiros adequados à educação infantil está fortemente associada à existência de outros recursos estruturais, como parques infantis e equipamentos multimídia. Escolas que investem na qualidade desses espaços tendem a criar um ambiente mais adequado para a aprendizagem e mais confiável para os pais, aumentando a procura por matrículas. Se o gestor precisar fazer cortes orçamentários, deve evitar reduzir a qualidade da infraestrutura básica, pois isso pode afastar os responsáveis e diminuir o número de novas matrículas. Caso precise priorizar investimentos, os banheiros adequados e os parques infantis devem ser mantidos, pois possuem impacto direto na escolha das famílias.

Além disso, a presença de laboratórios de informática está relacionada à disponibilidade de equipamentos multimídia, sugerindo que escolas que adotam uma abordagem tecnológica fazem isso de forma integrada. Esse fator demonstra que, para que a introdução de novas tecnologias seja eficiente, não basta apenas adquirir computadores ou dispositivos multimídia isoladamente, mas garantir que o ambiente escolar esteja preparado para oferecer suporte adequado ao uso desses recursos. Se um gestor deseja modernizar a escola sem comprometer recursos essenciais, deve planejar a integração entre laboratórios e equipamentos multimídia, assegurando que ambos possam ser utilizados de forma complementar.

O número de turmas e docentes também exerce influência sobre a dinâmica do ensino infantil, mas sua relação com a infraestrutura merece atenção. Escolas que possuem um menor número de turmas tendem a contar com menos recursos estruturais, indicando que a capacidade de atendimento e a qualidade da infraestrutura escolar caminham juntas. Se houver necessidade de redução no número de turmas, é fundamental que isso seja feito sem comprometer a qualidade dos espaços físicos e dos recursos pedagógicos oferecidos. O gestor que precisar reduzir turmas deve, ao mesmo tempo, buscar estratégias para manter a infraestrutura em bom estado, pois a falta de recursos pode afetar a atratividade da escola e, consequentemente, reduzir ainda mais as matrículas ao longo do tempo.

No **ensino fundamental**, os fatores que influenciam a atratividade e permanência dos alunos incluem a disponibilidade de professores, a quantidade de turmas e a infraestrutura educacional. Além de afetar diretamente o número de matrículas, a quantidade de professores e turmas também se relaciona com a presença de funcionários administrativos. Escolas que possuem um

corpo técnico mais estruturado tendem a oferecer um ambiente de ensino mais organizado, o que pode impactar indiretamente a retenção dos alunos e o desempenho escolar. Isso indica que cortes no quadro de funcionários administrativos devem ser planejados com cautela, pois uma equipe reduzida pode sobrecarregar os professores e comprometer a qualidade da gestão escolar. Caso o gestor precise otimizar os recursos, deve priorizar a manutenção do quadro docente e, se necessário, redistribuir funções dentro da equipe administrativa para evitar a sobrecarga dos professores.

A presença de bibliotecas e salas de leitura demonstrou ter um impacto positivo na retenção dos alunos, mas sua eficácia é ampliada quando combinada com o uso de equipamentos multimídia. Escolas que contam com bibliotecas estruturadas e, ao mesmo tempo, possuem tecnologia de apoio, como retroprojetores e computadores, proporcionam um ambiente de aprendizado mais dinâmico. Isso sugere que, caso haja necessidade de cortes em tecnologia, as bibliotecas devem ser preservadas para minimizar os impactos sobre a qualidade do ensino. Se não for possível manter o investimento total em tecnologia, o gestor pode adotar medidas como a reorganização do uso de retroprojetores e a ampliação de horários de funcionamento das bibliotecas para compensar a possível perda de outros recursos.

Além disso, foi observado que a mera existência de salas de aula não é um fator isolado determinante para a taxa de matrículas. Para que novas salas de aula tenham um impacto positivo, elas devem ser acompanhadas de investimentos em professores e recursos pedagógicos que garantam um ensino de qualidade. Caso o gestor precise expandir a escola, deve garantir que a construção de novas salas seja acompanhada pela contratação de professores e pela aquisição de materiais didáticos. Se não houver um planejamento adequado, a ampliação física pode não gerar os resultados esperados e acabar gerando espaços subutilizados.

No **ensino médio**, a retenção dos alunos está fortemente associada à presença de laboratórios de ciências, infraestrutura esportiva e acesso à tecnologia. A existência de laboratórios permite um ensino mais dinâmico, principalmente em disciplinas como física, química e biologia, o que pode tornar o aprendizado mais atrativo e reduzir as taxas de evasão escolar. No entanto, a eficácia desses espaços depende da disponibilidade de professores capacitados e de uma infraestrutura tecnológica que permita a utilização plena desses ambientes. Isso indica que, se um gestor deseja tornar o ensino médio mais envolvente sem necessariamente aumentar o quadro de docentes, o investimento em laboratórios pode ser uma solução estratégica para melhorar a qualidade do aprendizado. Caso não haja recursos para expandir todos os laboratórios, o gestor pode focar na otimização dos espaços já existentes, garantindo que sejam utilizados por diferentes turmas e disciplinas de forma eficiente.

O acesso à tecnologia também demonstrou ser um fator relevante para a permanência dos alunos. A quantidade de computadores por aluno e a disponibilidade geral de computadores na escola afetam a forma como os estudantes se relacionam com o aprendizado. Escolas que oferecem melhores condições tecnológicas tendem a ter maior engajamento por parte dos alunos, o que reforça a importância de manter esses equipamentos disponíveis. Caso haja necessidade

de ajustes orçamentários, uma alternativa seria a reorganização do uso dos computadores em horários compartilhados, evitando uma redução drástica no acesso à tecnologia. Se a escola precisar reduzir o investimento em equipamentos, é recomendável buscar parcerias externas ou iniciativas governamentais que possam suprir essa demanda sem comprometer o ensino.

A infraestrutura esportiva, representada pela presença de quadras poliesportivas, demonstrou ter um papel significativo na retenção dos alunos. Atividades extracurriculares, como esportes, contribuem para o engajamento dos estudantes e para a construção de um ambiente escolar mais atrativo. A relação entre educação e práticas esportivas sugere que cortes na infraestrutura esportiva podem afetar indiretamente o desempenho e a permanência dos alunos. Se não for possível expandir as instalações esportivas, uma alternativa seria incentivar parcerias com espaços comunitários ou ampliar o uso de atividades extracurriculares dentro da própria escola. Se o gestor precisar priorizar investimentos, manter quadras em bom estado e promover eventos esportivos pode ser uma estratégia para manter os alunos engajados sem necessidade de grandes gastos estruturais.

Diante desses achados, fica evidente que a gestão escolar deve considerar não apenas os fatores que impactam diretamente as matrículas, mas também as conexões internas entre os diferentes aspectos da escola. Se houver necessidade de reduzir custos, a prioridade deve ser a preservação do quadro docente, da infraestrutura básica e dos espaços pedagógicos que tenham impacto comprovado na retenção e aprendizado dos alunos. Construir novas salas sem garantir professores suficientes ou cortar investimentos em laboratórios sem compensar com outras estratégias podem gerar impactos negativos a longo prazo.

Assim, qualquer tomada de decisão deve considerar não apenas o impacto imediato, mas também os efeitos indiretos que determinadas mudanças podem causar no funcionamento geral da escola. O equilíbrio entre a infraestrutura, o corpo docente e os recursos tecnológicos deve ser mantido para garantir um ambiente escolar funcional e atrativo, permitindo que os alunos tenham acesso a um ensino de qualidade e que a escola mantenha sua capacidade de reter estudantes ao longo dos anos.

5

Conclusão

Este estudo teve como objetivo analisar os fatores que influenciam a quantidade de matrículas nos ensinos infantil, fundamental e médio na rede pública de Maceió, utilizando redes neurais LSTM para modelagem de séries temporais e técnicas de *Explainable AI*, como SHAP e Causalidade de Granger, para interpretar as variáveis mais relevantes. A partir dos dados do Censo Escolar do INEP, cobrindo um período de dez anos, foi possível identificar padrões que auxiliam na compreensão da dinâmica das matrículas e fornecem subsídios para uma gestão educacional mais eficiente.

Os resultados sugerem que fatores como infraestrutura escolar, disponibilidade de professores, acesso a tecnologia e oferta de atividades complementares desempenham papéis importantes na decisão dos alunos e suas famílias de permanecerem na escola. Além disso, identificou-se interdependência entre variáveis estruturais e pedagógicas, reforçando a necessidade de um planejamento integrado para evitar cortes orçamentários que possam impactar negativamente a retenção de alunos.

Apesar dos avanços obtidos, o estudo apresenta limitações. A análise restringiu-se à cidade de Maceió, o que limita a generalização dos achados para outras regiões. Além disso, a escolha da arquitetura LSTM, embora eficaz, poderia ser comparada com outros modelos de aprendizado de máquina para avaliar diferenças de desempenho.

Dessa forma, recomenda-se que pesquisas futuras ampliem a análise para outras cidades de Alagoas e estados do Brasil, considerando um conjunto de dados mais abrangente. Também seria interessante incorporar novas variáveis socioeconômicas e demográficas, a fim de refinar ainda mais as previsões. Além disso, o desenvolvimento de sistemas preditivos interativos baseados em dashboards educacionais pode auxiliar gestores na tomada de decisões estratégicas em tempo real, promovendo uma gestão mais dinâmica e orientada por dados.

Por fim, este estudo reforça a importância do uso de modelos preditivos e inteligência artificial explicável na educação, demonstrando que abordagens baseadas em aprendizado de má-

quina podem contribuir significativamente para o planejamento educacional, auxiliando na otimização de recursos e na formulação de políticas públicas mais eficazes.

Referências bibliográficas

[Alloghani et al. 2020]ALLOGHANI, M. et al. A systematic review on supervised and unsupervised machine learning algorithms for data science. In: BERRY, M. W.; MOHAMED, A.; YAP, B. W. (Ed.). *Supervised and Unsupervised Learning for Data Science: Unsupervised and Semi-Supervised Learning*. [S.l.]: Springer, 2020.

[Bakhtierzhon e Petrusevich 2024]BAKHTIERZHON, P.; PETRUSEVICH, D. Neural network analysis in time series forecasting. *Russian Technological Journal*, v. 12, n. 4, p. 106–116, 2024.

[Barros 2003]BARROS, M. *Séries Temporais e Modelagem Estatística*. 2003. Site da M. Barros Consultoria Ltda. 150p.

[Bengio e Simard 2019]BENGIO, Y.; SIMARD, P. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, v. 5, n. 2, p. 157–166, mar 2019.

[Bochie et al. 2020]BOCHIE, K. et al. Aprendizado profundo em redes desafiadoras: Conceitos e aplicações. *Sociedade Brasileira de Computação*, 2020.

[Brownlee 2019]BROWNLEE, J. *How to Fix the Vanishing Gradients Problem Using the ReLU*. 2019. Machine Learning Mastery.

[Christoper 2015]CHRISTOPER, O. *Understanding LSTM Networks*. 2015. Colah. Disponível em: https://colah.github.io/posts/2015-08-Understanding-LSTMs/.

[Dantas e Weydmann 2009]DANTAS, F.; WEYDMANN, C. L. Carne de frango: uma análise da relação entre os preços dos produtores e de exportação. *Revista de Economia e Agronegócio*, v. 7, n. 1, p. 31–53, 2009.

[Dickey e Fuller 1981]DICKEY, D. A.; FULLER, W. A. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica: Journal of the Econometric Society*, v. 49, n. 4, p. 1057–1072, 1981.

[EPSJV, Fiocruz 2024]EPSJV, Fiocruz. *Censo Escolar revela queda de 150 mil matrículas no Ensino Médio em 2023*. 2024. Disponível em: <https://www.epsjv.fiocruz.br/noticias/reportagem/censo-escolar-revela-queda-de-150-mil-matriculas-no-ensino-medio-em-2023>.

[Farias e Sáfadi 2010]FARIAS, H. P.; SáFADI, T. Causalidade entre as principais bolsas de valores do mundo. *RAM. Revista de Administração Mackenzie*, v. 11, n. 2, p. 96–122, 2010.

[Goodfellow e Bengio 2016]GOODFELLOW, I.; BENGIO. *Deep Learning*. [S.l.]: The Mit Press, 2016.

[Granger 1969]GRANGER, C. W. J. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, v. 37, n. 3, p. 424–438, 1969.

[Géron 2019]GéRON, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. 2. ed. Sebastopol, CA, USA: O'Reilly Media, 2019.

[Hochreiter e Schmidhuber 1997]HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. *Neural Computation*, v. 9, n. 8, p. 1735–1780, 1997.

[Josef 1991]JOSEF, H. *Untersuchungen zu dynamischen neuronalen Netzen*. Dissertação (Mestrado) — Technische Universität München, jun 1991. 15 jun. 1991.

[Kingma e Ba 2014]KINGMA, D.; BA, J. Adam: A method for stochastic optimization. *arXiv preprint*, arXiv:1412.6980, 2014.

[Kirsten 2019]KIRSTEN, G. *A evolução da descida de gradiente e seus otimizadores*. 2019. Medium. Disponível em: https://medium.com/@gabrielkirsten/a-evolu%C3%A7%C3%A3o-da-descida-de-gradiente-e-seus-otimizadores-680c835c1b4f.

[Le et al. 2019]LE, X. H. et al. Application of long short-term memory (lstm) neural network for flood forecasting. *Water (Switzerland)*, v. 11, n. 7, 2019.

[Lima e Sousa 2014]LIMA, A. A. A. A. d.; SOUSA, F. P. d. *Censo Escolar da Educação Básica: Uma Referência para Elaboração de Políticas Públicas e Transferência de Recursos para Educação Pública*. 2014. Revista Com Censo, Vol. 1, n.1, 1ª ed., Brasília-DF, p.p. 94-102, dezembro de 2014. Disponível em: http://www.cre.se.df.gov.br/ascom/documentos/suplav/revista_comcenso/artigo_censo_escolar_da_ed_basica.pdf.

[Lundberg e Lee 2017]LUNDBERG, S. M.; LEE, S.-I. A unified approach to interpreting model predictions. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS)*. [S.l.]: Curran Associates Inc., 2017. p. 4765–4774.

[Lv et al. 2022]LV, L. et al. A vmd and lstm based hybrid model of load forecasting for power grid security. *IEEE Transactions on Industrial Informatics*, v. 18, n. 9, p. 6474–6482, September 2022.

[Morettin 1987]MORETTIN, P. A. *Previsão de series temporais*. São Paulo: Atual, 1987.

[Ni et al. 2020]NI, L. et al. Streamflow and rainfall forecasting by two long short-term memory-based models. *Journal of Hydrology*, v. 583, April 2020.

[Phillips e Ouliaris 1990]PHILLIPS, P. C. B.; OULIARIS, S. Asymptotic properties of residual based tests for cointegration. *Econometrica: Journal of the Econometric Society*, v. 58, n. 1, p. 165–193, 1990.

[Phillips e Perron 1988]PHILLIPS, P. C. B.; PERRON, P. Testing for a unit root in time series regression. *Biometrika*, v. 75, n. 2, p. 335–346, 1988.

[Souza e Oliveira 2012]SOUZA, R. M. d. O.; OLIVEIRA, E. A. M. O censo escolar no contexto da democratização da educação básica e do pacto federativo brasileiro. In: JUNQUEIRA&MARIN EDITORES, CAMPINAS. *XVI ENDIPE - Encontro Nacional de Didática e Práticas de Ensino*. [S.l.], 2012. Livro 3.

[TATU 2021]TATU, A. *Maceió é a capital com maior abandono escolar do país, aponta painel*. [S.l.]: Agência Tatu de Jornalismo de Dados, 2021. Disponível em: https://www.agenciatatu.com.br/noticia/maceio-e-a-capital-com-maior-abandono-escolar-do-pais-aponta-painel/.

[Toda e Yamamoto 1995]TODA, H. Y.; YAMAMOTO, T. Statistical inference in vector autoregressions with possibly integrated processes. *Journal of Econometrics*, v. 66, n. 1-2, p. 225–250, 1995.

[UNICEF 2020]UNICEF. *O mapa da exclusão escolar em Maceió*. Brasília: UNICEF, 2020. Disponível em: https://www.unicef.org/brazil/relatorios/plataforma-dos-centros-urbanos-2017-2020/mapa-exclusao-escolar-maceio.

[Velden 2024]VELDEN, B. H. M. Van der. Explainable ai: current status and future potential. *European Radiology*, v. 34, p. 1187–1189, 2024.

[Yegulalp 2019]YEGULALP, S. *What is TensorFlow? The machine learning library explained*. 2019. InfoWorld. Disponível em: https://www.infoworld.com/article/3278008/what-is-tensorflow-the-machine-learning-library-explained.html.