

Análise e Previsão da Evasão Escolar na Educação Infantil: Uma Abordagem de Séries Temporais com Foco em Políticas Públicas

Ana Cláudia de Lima Aleixo¹, Ana Carolina Martins²,

Eduardo David³, Gustavo Santiago⁴

¹Faculdade de Computação e Informática (FCI)
Universidade Presbiteriana Mackenzie – São Paulo, SP – Brasil

10414992@mackenzie.br,

10174844@mackenzie.br,

10414643@mackenzie.br,

10414643@mackenzie.br.

Resumo: Este projeto tem como objetivo analisar e prever a evasão escolar na Educação Infantil, com foco em creches e pré-escolas da rede pública brasileira. A pesquisa considera recortes por tipo de turno (integral ou parcial), rede de ensino (estadual ou municipal) e localização (urbana ou rural), a fim de compreender padrões regionais e identificar cenários críticos. A partir de dados do Censo Escolar do INEP (2015–2023), o estudo utiliza técnicas de séries temporais combinadas com algoritmos de aprendizado de máquina para realizar previsões e apoiar a formulação de políticas públicas voltadas à permanência escolar. A análise considera que fatores estruturais e socioeconômicos influenciam fortemente a evasão, especialmente em contextos de vulnerabilidade. Como contribuição extensionista, o projeto se alinha ao ODS 4 da ONU, promovendo educação inclusiva e de qualidade. Espera-se que os resultados possam orientar gestores públicos na implementação de ações mais eficazes e baseadas em evidências, além de fomentar o uso de tecnologia e ciência de dados na educação pública brasileira.

Palavras-chave: Evasão Escolar, Educação Infantil, Séries Temporais

Abstract: This project aims to analyze and forecast school dropout rates in Early Childhood Education, focusing on public daycare and preschool institutions in Brazil. The study segments data by school shift (full-time or part-time), school network (state or municipal), and location (urban or rural), in order to understand regional patterns and identify critical dropout scenarios. Using data from the INEP School Census (2015–2023), the project applies time series techniques combined with machine learning algorithms to make

predictions and support public policies that promote school retention. The analysis acknowledges that structural and socioeconomic factors strongly influence dropout rates, especially in vulnerable contexts. As a community-oriented initiative, the project aligns with the UN's SDG 4, promoting inclusive and quality education. The expected outcome is to provide actionable insights for public education managers and encourage the adoption of data science tools in educational planning and decision-making.

Keywords: *School Dropout, Early Childhood Education, Time Series*

SUMÁRIO

1. Introdução	4
2. Objetivo	6
3. Descrição da Base de Dados.....	6
4. Referencial Teórico.....	7
5. Pipeline da Solução	8
6. Cronograma do Projeto.....	10
8. Referências bibliográficas	10

1. Introdução

A evasão escolar na Educação Infantil representa um desafio estrutural para o sistema educacional brasileiro, pois impacta diretamente o desenvolvimento integral das crianças e compromete sua permanência nas etapas seguintes da educação básica. A Educação Infantil — composta por creches e pré-escolas — é a base para o desenvolvimento cognitivo, afetivo e social. No entanto, estudos mostram que crianças de contextos vulneráveis, especialmente nas redes públicas e áreas rurais, têm maior risco de interrupção precoce da trajetória escolar (Silva & Andrade, 2020).

O Censo Escolar, realizado anualmente pelo INEP, apresenta dados sobre matrículas, fluxo e rendimento dos estudantes da Educação Básica no Brasil. Segundo o relatório de 2023, "embora haja avanços na ampliação do acesso, as taxas de abandono permanecem expressivas em determinados contextos, especialmente na educação infantil em tempo parcial e em áreas rurais" (INEP, 2023, p. 14). Isso revela a urgência de análises segmentadas e preditivas que possam embasar decisões políticas mais precisas e eficazes.

A dificuldade em acompanhar a evasão escolar com granularidade suficiente — como por etapa de ensino, turno e rede — limita a capacidade de intervenção estratégica dos gestores públicos. Além disso, a maior parte dos estudos existentes se concentra no Ensino Fundamental e Médio, deixando a Educação Infantil em segundo plano. No entanto, como destacam Costa e Lima (2021), “os primeiros anos de vida são determinantes para a formação de habilidades cognitivas e emocionais que sustentam o sucesso escolar futuro” (p. 89). A ausência de políticas focalizadas neste estágio compromete não só o direito à educação, mas o desenvolvimento humano em longo prazo.

O uso de séries temporais combinadas com técnicas de aprendizado de máquina tem se mostrado eficaz para entender padrões educacionais e antecipar cenários críticos. Modelos como Prophet e LSTM têm sido empregados em estudos para prever matrículas, evasão e desempenho escolar, permitindo respostas mais rápidas e eficientes por parte dos órgãos responsáveis (Ferreira et al., 2022). A proposta deste projeto, portanto, é não apenas identificar tendências históricas, mas criar um produto analítico que possa ser reutilizado e adaptado por secretarias de educação, ONGs e gestores escolares para promover permanência, equidade e qualidade na Educação Infantil.

Neste contexto, este projeto busca analisar séries temporais relacionadas à evasão escolar na Educação Infantil, considerando a segmentação por tipo de turno (parcial ou

integral), rede de ensino (estadual ou municipal) e localização (urbana ou rural). O objetivo é aplicar modelos de aprendizado de máquina para prever padrões futuros e apoiar políticas públicas baseadas em evidências. A análise contribuirá para os Objetivos de Desenvolvimento Sustentável, em especial o ODS 4, que visa "assegurar a educação inclusiva, equitativa e de qualidade" até 2030 (ONU, 2015).

A evasão escolar na Educação Infantil é um fenômeno silencioso e, muitas vezes, negligenciado pelas políticas públicas educacionais. Embora os dados do Censo Escolar evidenciem avanços no acesso à creche e à pré-escola, a permanência e a continuidade da criança no ambiente educacional ainda são desafiadoras, especialmente em contextos vulneráveis. A ausência de dados consolidados e análises preditivas nesse segmento limita a ação do Estado e fragiliza o direito à educação na primeira infância, fase determinante para o desenvolvimento humano.

Este projeto nasce da percepção de que a evasão escolar, quando diagnosticada precocemente, pode ser combatida com políticas públicas assertivas, baseadas em evidências e orientadas por dados. A aplicação de modelos de séries temporais permite não apenas compreender a trajetória histórica do problema, mas antecipar cenários críticos e agir preventivamente. Essa abordagem oferece uma vantagem estratégica para gestores públicos, ao possibilitar decisões fundamentadas em projeções reais e segmentadas por rede, turno e localização.

A relevância do tema também está relacionada ao seu impacto social e à necessidade de romper ciclos de desigualdade. Crianças que abandonam a escola precocemente tendem a apresentar maiores dificuldades acadêmicas ao longo da vida, menor inserção no mercado de trabalho e menor acesso a oportunidades. Investir em estratégias de permanência desde a Educação Infantil é, portanto, investir em equidade, inclusão e desenvolvimento sustentável, pilares fundamentais do ODS 4 — Educação de Qualidade.

Além disso, a proposta do projeto atende ao caráter extensionista exigido pela formação acadêmica, ao buscar devolver à sociedade um produto analítico acessível, útil e de aplicação prática. A construção de um modelo preditivo com base em dados públicos primários fortalece a transparência, o controle social e o uso da ciência de dados em benefício do coletivo. Dessa forma, o projeto propõe não apenas uma solução técnica, mas uma ferramenta de transformação social que contribui com a formulação de políticas educacionais mais justas e eficazes.

2. Objetivo

O principal objetivo deste projeto é desenvolver um produto analítico capaz de analisar, interpretar e prever a evasão escolar na Educação Infantil, com base nos históricos do Censo Escolar organizados em séries temporais. A proposta envolve a segmentação dos dados por tipo de turno (parcial ou integral), rede de ensino (estadual ou municipal) e localização geográfica (urbana ou rural), de modo a identificar padrões e comportamentos distintos entre os diferentes contextos educacionais. Com isso, espera-se compreender a evolução da evasão ao longo do tempo e antecipar tendências futuras.

Como metas específicas, o projeto visa: (i) estruturar e limpar os dados educacionais públicos de forma adequada para análise temporal; (ii) aplicar modelos estatísticos e algoritmos de aprendizado de máquina — como Prophet ou LSTM — para geração de previsões; (iii) produzir visualizações e indicadores acessíveis para o público interessado, especialmente gestores educacionais; e (iv) disponibilizar publicamente o produto final como contribuição extensionista alinhada aos Objetivos de Desenvolvimento Sustentável, com destaque para o ODS 4. A intenção é que esse projeto sirva como base para decisões mais informadas, políticas públicas mais eficazes e ações concretas de enfrentamento à evasão escolar desde os primeiros anos da educação básica.

3. Descrição da Base de Dados

O conjunto de dados utilizado neste projeto foi extraído do Censo Escolar da Educação Básica, disponível no portal do INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira). Trata-se de uma base oficial e de fonte primária, composta por dados anuais e declaratórios fornecidos pelas instituições de ensino públicas e privadas de todo o país. Para este estudo, foram utilizados os dados referentes ao período de 2020 a 2024, com foco específico nas etapas da Educação Infantil, contemplando Creche e Pré-escola.

A base contém informações detalhadas sobre o número de matrículas por escola, etapa de ensino, turno (parcial ou integral), rede administrativa (estadual ou municipal) e localização (urbana ou rural). Além disso, inclui indicadores relevantes como taxa de abandono, situação de matrícula, permanência, e códigos territoriais que permitem a agregação dos dados por região, município ou estado. A estrutura dos arquivos segue o

formato CSV (Comma Separated Values), com milhares de registros por ano, organizados em campos padronizados e dicionários de variáveis disponibilizados pelo próprio INEP.

A preparação dos dados exigiu etapas de filtragem, categorização e reestruturação para permitir a análise temporal. Foram selecionadas apenas as instituições públicas de ensino, e os dados foram organizados em séries temporais com granularidade anual. Essa organização possibilita a modelagem estatística e a aplicação de algoritmos de previsão, mantendo a rastreabilidade e a integridade das informações. A escolha dessa base garante confiabilidade, atualidade e alinhamento com os objetivos extensionistas do projeto.

4. Referencial Teórico

A evasão escolar é um dos principais desafios da educação pública brasileira e pode ser compreendida como o abandono da trajetória escolar antes da conclusão da etapa educacional correspondente. Segundo Cunha (2020), "o fenômeno da evasão é multifacetado, envolvendo fatores socioeconômicos, culturais e estruturais que ultrapassam os muros da escola" (p. 112). Esses fatores variam conforme o território, o nível de ensino e o perfil da população atendida.

Apesar de sua importância, a Educação Infantil enfrenta sérios desafios em relação ao acesso, permanência e qualidade, especialmente nas redes públicas. Segundo dados do IBGE (2022), ainda há uma considerável desigualdade no atendimento entre áreas urbanas e rurais, bem como entre redes estaduais e municipais. Essa disparidade reflete não apenas a disponibilidade de vagas, mas também as condições estruturais e pedagógicas das instituições de ensino infantil.

Séries temporais são sequências de dados ordenadas no tempo, geralmente coletadas em intervalos regulares, e são amplamente utilizadas para análise de tendências, sazonalidades e previsões em diversas áreas, incluindo a educação. De acordo com Hyndman e Athanasopoulos (2021), "o uso de séries temporais permite a modelagem de padrões históricos para projetar cenários futuros e orientar decisões estratégicas" (p. 15). Esse tipo de análise é particularmente útil para políticas públicas, pois permite antecipar demandas e prevenir problemas estruturais.

Na educação, a aplicação de modelos de séries temporais tem ganhado espaço por sua capacidade de prever fenômenos como matrículas, taxas de evasão, desempenho

escolar e até alocação de recursos. Modelos estatísticos tradicionais como ARIMA vêm sendo amplamente utilizados, mas, mais recentemente, algoritmos de aprendizado de máquina como LSTM e Prophet têm se mostrado eficazes em contextos educacionais (Ferreira et al., 2022). Essas abordagens oferecem maior flexibilidade na captura de padrões não lineares e de múltiplas variáveis.

No presente projeto, o uso de séries temporais permitirá identificar os momentos de maior risco de evasão, a eficácia de políticas passadas e a projeção de cenários futuros. Ao integrar dados segmentados por rede, tipo de turno e localização, o modelo possibilita uma visão mais rica e útil para a gestão pública. Como destacam Souza e Lima (2020), “a análise temporal em educação ainda é subexplorada, mas tem um potencial imenso para apoiar decisões mais inteligentes e eficazes” (p. 102).

5. Pipeline da Solução

A solução proposta para o problema da evasão escolar na Educação Infantil está organizada em um pipeline dividido em sete etapas principais. Cada fase foi planejada para garantir a qualidade dos dados, a eficácia da análise e a aplicabilidade prática dos resultados. A seguir, detalha-se o pipeline da solução, com a respectiva discussão de suas funções e relevância no contexto do projeto.

Coleta de Dados

A primeira etapa consiste na obtenção dos microdados do Censo Escolar da Educação Básica diretamente do site do INEP, no formato original CSV. Os dados coletados referem-se ao período de 2020 a 2024, com foco exclusivo em instituições públicas de Educação Infantil. A escolha da fonte primária garante a confiabilidade e atualidade das informações, em conformidade com as diretrizes metodológicas do projeto.

Pré-processamento dos Dados

Essa etapa é essencial para transformar os dados brutos em uma estrutura adequada para análise temporal. Inclui o filtro de registros relevantes (creches e pré-escolas), remoção de inconsistências e dados ausentes, além da categorização por rede de ensino (estadual/municipal), localização (urbana/rural) e turno (parcial/integral). Ao final

dessa etapa, os dados são reestruturados para formar séries temporais anuais agregadas por segmento educacional.

Análise Exploratória dos Dados (EDA)

A EDA tem como objetivo identificar padrões, tendências e anomalias nos dados. Serão construídos gráficos de linha, barras e mapas para visualizar o comportamento da evasão escolar ao longo do tempo e entre diferentes categorias. Essa fase é importante não apenas para gerar hipóteses sobre os fatores envolvidos, mas também para direcionar a escolha dos modelos de previsão mais adequados.

Modelagem Preditiva

Nesta etapa, serão aplicados modelos de séries temporais, com destaque para algoritmos estatísticos (ARIMA, Prophet) e técnicas de aprendizado de máquina (LSTM, se viável). Os modelos serão treinados com os dados de 2020 a 2023 e validados com o ano de 2024. O objetivo é prever as taxas de evasão escolar nos anos seguintes, possibilitando a identificação antecipada de contextos críticos.

Avaliação e Validação dos Modelos

Serão utilizadas métricas como RMSE (Root Mean Squared Error), MAE (Mean Absolute Error) e MAPE (Mean Absolute Percentage Error) para avaliar o desempenho dos modelos. A escolha do modelo final se dará com base na acurácia preditiva e na capacidade de gerar insights interpretáveis para gestores educacionais.

Visualização e Comunicação dos Resultados

Será desenvolvido um painel analítico com visualizações claras, objetivas e acessíveis, com o intuito de facilitar a tomada de decisão por parte de órgãos públicos, pesquisadores e instituições educacionais. Os dashboards incluirão gráficos temporais, comparações por rede e turno, e mapas de calor por região, possibilitando uma leitura intuitiva dos dados.

Produto e Divulgação Pública

Por fim, todo o código, base de dados processada e documentação serão disponibilizados em repositório público (GitHub), como forma de promover a

reprodutibilidade e o caráter extensionista do projeto. A solução será apresentada a partir de um relatório técnico e, se possível, por meio de eventos, seminários ou oficinas com gestores públicos, contribuindo para a implementação de políticas alinhadas ao ODS 4 (Educação de Qualidade).

6. Cronograma do Projeto

Entrega	Data
Definição do projeto e equipe	28/02
Referencial Teórico e Cronograma	28/03
Implementação Parcial	25/04
Implementação e Entrega Final	30/05

8. Referências Bibliográficas

CUNHA, Rogério da Silva. Panorama da evasão escolar no Brasil: causas e consequências. *Revista Brasileira de Educação*, v. 25, n. 81, p. 110-125, 2020.

FERREIRA, Lucas R.; SANTOS, Renata M.; OLIVEIRA, Marcos A. Modelos preditivos aplicados à evasão escolar: uma revisão sistemática. *Cadernos de Informática*, v. 10, n. 2, p. 45-63, 2022.

HECKMAN, James J. The Economics of Inequality: The Value of Early Childhood Education. *American Educator*, v. 35, n. 1, p. 31-47, 2011.

HYNDMAN, Rob J.; ATHANASOPOULOS, George. *Forecasting: Principles and Practice*. 3. ed. Melbourne: OTexts, 2021. Disponível em: <https://otexts.com/fpp3/>. Acesso em: 26 mar. 2025.

IBGE. Educação 2022: Pesquisa Nacional por Amostra de Domicílios Contínua - PNAD. Rio de Janeiro: IBGE, 2022. Disponível em: <https://www.ibge.gov.br/>. Acesso em: 26 mar. 2025.

INEP. Censo Escolar da Educação Básica 2023: resumo técnico. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, Brasília: INEP, 2023.

OLIVEIRA, Cláudia M.; MENDES, Patrícia A. Importância da Educação Infantil no desenvolvimento integral da criança. *Revista Educação em Foco*, v. 22, n. 1, p. 30-41, 2019.

RIBEIRO, Tânia M.; SILVA, Felipe C. Motivos da evasão escolar no Brasil: um olhar sociológico. *Revista Educação e Sociedade*, v. 42, n. 154, p. 72-84, 2021.

SOUZA, Camila R.; LIMA, Eduardo H. Análises temporais aplicadas à educação pública brasileira: tendências e desafios. *Revista Brasileira de Ciências Aplicadas*, v. 12, n. 3, p. 95-105, 2020.

Grossi, et al. (2021). Editorial: Greenhouse Gas Emissions and Terrestrial Ecosystems. *Frontiers*.

Kumar, et al. (2021). Metodologia holística para estimativa de emissões e remoções anuais de GEE em parques naturais. *Frontiers in Environmental Science*

Grossi, et al. (2021). "Métodos para estimar emissões e remoções de GEE." *Frontiers in Environmental Science*.

Hui, et al. (2021). "Impacto das atividades humanas nas emissões de GEE." *Journal of Environmental Management*.

Kumar, et al. (2021). "Conservação de terra e gestão de carbono." *Science Advances*.