v7.n12.2026

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

ANÁLISE PREDITIVA NA GESTÃO EDUCACIONAL: UMA FERRAMENTA PARA ESTIMATIVA DA EVASÃO NO ENSINO SUPERIOR

PREDICTIVE ANALYTICS IN EDUCATIONAL MANAGEMENT: A TOOL FOR ESTIMATING STUDENT DROPOUT IN HIGHER EDUCATION

ANÁLISIS PREDICTIVO EN LA GESTIÓN EDUCATIVA: UNA HERRAMIENTA PARA LA ESTIMACIÓN DE LA DESERCIÓN EN LA EDUCACIÓN SUPERIOR

Barbara A. de Araujo[1], Felipe Diniz Dallilo[2], Fabiana Florian[3]

https://doi.org/10.47820/recima21.v7i12.9880

PUBLICADO: 12/2026

RESUMO

A evasão no Ensino Superior afeta indicadores acadêmicos, financeiros e administrativos das instituições. Este trabalho desenvolveu uma ferramenta computacional de execução local que utiliza os microdados públicos do Censo da Educação Superior (INEP) para estimar a taxa de evasão de cursos de graduação no ano seguinte, por meio do algoritmo *Random Forest Regressor*. A taxa de evasão foi operacionalizada como a razão entre os alunos desvinculados e o total de alunos com vínculo ao curso no ano censitário, e o problema foi formulado de modo que os atributos de um ano estimem a evasão do ano subsequente, evitando o vazamento de dados. Com base nos cursos de bacharelado presenciais do estado de São Paulo nas edições de 2022 a 2024, o modelo foi treinado com o par 2022→2023 (4.223 cursos) e avaliado no par 2023→2024 (4.156 cursos). O *Random Forest* obteve erro médio absoluto de 6,53 pontos percentuais e R^2 de 0,248, reduzindo o erro em 31,5% em relação à previsão pela média e em 9,9% em relação à repetição da taxa do ano anterior. Entre os cursos sinalizados com risco alto, 79,5% registraram de fato evasão igual ou superior a 20%. A taxa de evasão do ano corrente, a taxa de trancamento e a categoria administrativa da instituição foram os atributos mais relevantes. Os resultados indicam que dados públicos padronizados permitem construir estimativas úteis ao planejamento institucional, embora a variação anual da evasão de cada curso seja apenas parcialmente previsível.

PALAVRAS-CHAVE: Evasão. Ensino Superior. Aprendizado de Máquina. *Random Forest*. Censo da Educação Superior.

ABSTRACT

Student dropout in Higher Education affects the academic, financial and administrative indicators of institutions. This work developed a locally executed software tool that uses the public microdata of the Brazilian Higher Education Census (INEP) to estimate the following year's dropout rate of undergraduate programs using the Random Forest Regressor algorithm. The dropout rate was defined as the ratio between students who left the program and all students enrolled in it during the census year, and the problem was formulated so that the attributes of one year estimate the dropout rate of the next, avoiding data leakage. Using in-person bachelor's programs in the state of São Paulo from the 2022 to 2024 editions, the model was trained on the 2022→2023 pairs (4,223 programs) and evaluated on the 2023→2024 pairs (4,156 programs).

[1] Graduanda em Sistemas de Informação pela Universidade de Araraquara (UNIARA).
[2] Doutor em Engenharia de Software pela Universidade de São Paulo (USP) e docente na Universidade de Araraquara (UNIARA).
[3] Doutora em Alimentos e Nutrição pela Universidade Estadual Paulista (UNESP) e docente na Universidade de Araraquara (UNIARA).

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

The Random Forest achieved a mean absolute error of 6.53 percentage points and an R² of 0.248, reducing the error by 31.5% compared with predicting the mean and by 9.9% compared with repeating the previous year's rate. Among programs flagged as high risk, 79.5% actually had dropout rates of 20% or more. The current dropout rate, the enrollment suspension rate and the institution's administrative category were the most relevant attributes. The results show that standardized public data can support useful estimates for institutional planning, although the yearly variation in each program's dropout rate is only partially predictable.

KEYWORDS: *Student dropout. Higher Education. Machine Learning. Random Forest. Higher Education Census.*

RESUMEN

La deserción en la Educación Superior afecta los indicadores académicos, financieros y administrativos de las instituciones. Este trabajo desarrolló una herramienta de ejecución local que utiliza los microdatos públicos del Censo de la Educación Superior de Brasil (INEP) para estimar, con el algoritmo Random Forest Regressor, la tasa de deserción de los cursos de grado en el año siguiente. La tasa se definió como la razón entre los alumnos desvinculados y el total de alumnos con vínculo al curso en el año censal, y los atributos de un año se usaron para estimar la deserción del año siguiente, evitando la fuga de datos. Con los cursos presenciales de grado (bacharelado) del estado de São Paulo de 2022 a 2024, el modelo fue entrenado con el par 2022→2023 (4.223 cursos) y evaluado en el par 2023→2024 (4.156 cursos). El Random Forest obtuvo un error absoluto medio de 6,53 puntos porcentuales y un R² de 0,248, y redujo el error en un 31,5% frente a la predicción por la media y en un 9,9% frente a la repetición de la tasa del año anterior. El 79,5% de los cursos señalados con riesgo alto registró una deserción igual o superior al 20%. La tasa de deserción del año en curso, la tasa de suspensión de matrícula y la categoría administrativa fueron los atributos más relevantes. Los resultados indican que los datos públicos permiten estimaciones útiles para la planificación institucional, aunque la variación anual de la deserción sea solo parcialmente previsible.

PALABRAS CLAVE: *Deserción estudiantil. Educación Superior. Aprendizaje automático. Random Forest. Censo de la Educación Superior.*

1. INTRODUÇÃO

O presente trabalho propõe a aplicação de técnicas de análise preditiva na gestão educacional, com foco no desenvolvimento de uma ferramenta computacional capaz de estimar a taxa de evasão acadêmica em cursos de graduação por meio de algoritmos de *Machine Learning*. Em um cenário de crescente complexidade na gestão do Ensino Superior, compreender os fatores associados à evasão estudantil tornou-se um desafio estratégico para as Instituições de Ensino Superior (IES), uma vez que esse fenômeno impacta diretamente indicadores acadêmicos, financeiros e administrativos. Diferentemente de abordagens que dependem de bases de dados institucionais restritas, esta pesquisa fundamenta-se na utilização dos microdados do Censo da Educação Superior, disponibilizados anualmente pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), permitindo a construção de uma solução aplicável a diferentes cursos e instituições do país.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

O objetivo geral deste trabalho é desenvolver uma ferramenta computacional de execução local e arquitetura genérica, baseada em técnicas de *Machine Learning*, capaz de utilizar dados públicos do Censo da Educação Superior para estimar a taxa de evasão em cursos de graduação e fornecer subsídios analíticos para apoio à tomada de decisão na gestão educacional.

Para alcançar esse objetivo, foram definidos os seguintes objetivos específicos: (i) processar os microdados do Censo da Educação Superior de 2022 a 2024 e definir um indicador anual de evasão por curso; (ii) construir atributos relativos a partir das variáveis do Censo e organizar os dados de forma que as características de um ano sejam usadas para prever a evasão do ano seguinte; (iii) treinar o algoritmo *Random Forest Regressor* e comparar seu desempenho com modelos de referência; (iv) identificar os atributos mais relevantes para a previsão; e (v) desenvolver uma interface que permita consultar e exportar os resultados.

A relevância do tema fundamenta-se na necessidade de ampliar o acesso a ferramentas de análise de dados e apoio à tomada de decisão no contexto educacional. Muitas Instituições de Ensino Superior não dispõem de equipes especializadas em Ciência de Dados ou de recursos financeiros para aquisição de soluções analíticas avançadas. Ao estruturar uma ferramenta de execução local baseada em dados públicos e gratuitos, este estudo busca oferecer um instrumento de apoio à gestão acadêmica capaz de identificar padrões relacionados à evasão e fornecer estimativas que auxiliem no planejamento institucional. Além disso, a pesquisa contribui para a aplicação prática de conceitos de Ciência de Dados, *Business Intelligence* e Aprendizado de Máquina no setor educacional.

O problema que orienta esta pesquisa é: como estruturar uma ferramenta preditiva baseada em *Machine Learning* que use os microdados do Censo da Educação Superior para estimar a taxa de evasão em cursos de graduação de forma confiável e escalável?

Parte-se da hipótese de que variáveis presentes nos microdados do INEP, tais como quantidade de matriculados, ingressantes, concluintes, indicadores de financiamento estudantil, características institucionais e informações sociodemográficas dos estudantes, apresentam associação com os índices de evasão observados nos cursos de graduação. Dessa forma, espera-se que a aplicação de técnicas de engenharia de atributos e algoritmos de aprendizado supervisionado permita construir modelos capazes de estimar taxas de evasão com níveis satisfatórios de precisão e capacidade de generalização para diferentes cenários educacionais.

A pesquisa caracteriza-se como aplicada, explicativa e de abordagem quantitativa. Quanto aos procedimentos, combina pesquisa documental, para obtenção e tratamento dos microdados do Censo, e estudo experimental, para desenvolvimento, treinamento e validação

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

do modelo preditivo. Foram analisados os cursos de bacharelado presenciais do estado de São Paulo nas edições de 2022, 2023 e 2024. O algoritmo *Random Forest Regressor* foi utilizado para estimar a taxa de evasão do ano seguinte a partir das características do ano corrente, e seu desempenho foi comparado ao de modelos de referência.

2. REFERENCIAL TEÓRICO

2.1. Ciência de dados e aprendizado de máquina

O crescimento da disponibilidade de dados digitais tem impulsionado a utilização de técnicas de Ciência de Dados para extração de conhecimento e apoio à tomada de decisão. A Ciência de Dados reúne métodos estatísticos, computacionais e analíticos capazes de transformar grandes volumes de dados em informações úteis para organizações públicas e privadas.

Dentro desse contexto, a análise preditiva destaca-se como uma abordagem voltada à identificação de padrões históricos para realização de previsões sobre eventos futuros. Segundo Kuhn e Johnson (2013), modelos preditivos permitem explorar relações existentes entre variáveis e utilizá-las para estimar comportamentos ainda não observados.

2.2. Machine Learning e aprendizado supervisionado

O Aprendizado de Máquina (*Machine Learning*) constitui uma subárea da Inteligência Artificial dedicada ao desenvolvimento de algoritmos capazes de aprender padrões a partir de dados históricos e utilizar esse conhecimento para realizar previsões ou apoiar processos de tomada de decisão (FACELI *et al.*, 2011).

Os métodos de aprendizado supervisionado utilizam exemplos previamente rotulados para construir modelos capazes de prever resultados em novos conjuntos de dados. Entre suas aplicações destacam-se problemas de classificação e regressão, amplamente empregados em cenários de análise preditiva (KUHN; JOHNSON, 2013).

Em problemas de regressão, o objetivo consiste em estimar valores numéricos contínuos, permitindo prever indicadores quantitativos com base em características observadas anteriormente. Essa abordagem é especialmente útil quando se deseja modelar taxas, índices e tendências futuras.

Na construção de modelos preditivos, é preciso evitar o vazamento de dados (*data leakage*), que ocorre quando o modelo tem acesso, no treinamento, a informações que não

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

estariam disponíveis no momento da previsão. Nesse caso, as métricas de desempenho ficam mais otimistas do que o modelo realmente é (KAUFMAN *et al.*, 2012).

2.3. Random Forest como modelo de regressão

Entre os algoritmos de aprendizado supervisionado disponíveis, o *Random Forest* destaca-se por sua robustez e capacidade de generalização. Proposto por Breiman (2001), o método baseia-se na construção de múltiplas árvores de decisão combinadas para produzir previsões mais estáveis e precisas.

Na modalidade de regressão, o *Random Forest Regressor* permite estimar valores contínuos a partir da combinação dos resultados gerados pelas árvores individuais. Essa estratégia reduz problemas de sobreajuste (*overfitting*) e melhora a capacidade de generalização do modelo.

Outra característica relevante é a possibilidade de mensurar a importância relativa das variáveis utilizadas durante o treinamento, permitindo identificar quais atributos apresentam maior influência sobre os resultados previstos.

Neste trabalho, a importância foi estimada por permutação, procedimento proposto pelo próprio Breiman (2001): os valores de um atributo são embaralhados na base de teste e mede-se o quanto o erro do modelo aumenta. Quanto maior o aumento, maior a dependência do modelo em relação àquele atributo. Por ser calculada sobre dados não utilizados no treinamento, essa medida é menos sujeita aos vieses da importância baseada na redução de impureza das árvores.

A avaliação do desempenho de modelos baseados em regressão exige o uso de métricas estatísticas distintas das abordagens de classificação. Para mensurar a precisão do *Random Forest Regressor* na estimativa das taxas de evasão, foram aplicadas as métricas de Erro Médio Absoluto (MAE), Erro Quadrático Médio (MSE), sua raiz (RMSE), expressa na mesma unidade da variável-alvo, e o Coeficiente de Determinação (R^2), permitindo analisar o desvio das estimativas em relação aos dados reais observados.

A adoção dessas métricas está alinhada às recomendações de Kuhn e Johnson (2013), que discutem a avaliação de modelos de regressão por meio de medidas de erro absoluto e quadrático.

2.4. Dados públicos e o Censo da Educação Superior

A disponibilização de dados abertos governamentais tem ampliado as possibilidades de desenvolvimento de soluções analíticas em diferentes áreas do conhecimento. No Brasil, o

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) disponibiliza anualmente os dados do Censo da Educação Superior.

Essa base reúne informações relacionadas às Instituições de Ensino Superior, cursos de graduação, matrículas, ingressantes, concluintes, modalidades de ensino e indicadores acadêmicos. Por serem padronizados e disponibilizados em âmbito nacional, esses dados permitem a realização de estudos comparativos e o desenvolvimento de modelos preditivos aplicáveis em diferentes contextos.

2.5. Machine Learning aplicado à educação

A aplicação de técnicas de *Machine Learning* no contexto educacional vem recebendo cada vez mais atenção na pesquisa acadêmica. A área conhecida como Mineração de Dados Educacionais (*Educational Data Mining*) busca utilizar dados gerados por instituições de ensino para identificar padrões, prever comportamentos e apoiar processos decisórios.

Diversos estudos têm empregado modelos preditivos para análise de desempenho acadêmico, identificação de fatores associados à evasão e estimativa de indicadores institucionais. Segundo Silvestri, Souza e Vieira (2025), a utilização de algoritmos de aprendizado de máquina possibilita extrair informações relevantes a partir de bases educacionais, contribuindo para o desenvolvimento de estratégias mais eficientes de gestão.

No Brasil, a produção científica sobre permanência e evasão na educação superior é extensa (MACIEL; CUNHA JÚNIOR; LIMA, 2019) e caracteriza-se pela diversidade de abordagens metodológicas e teóricas (STARK; PINTO, 2025), o que reforça a importância de explicitar como o fenômeno é medido em cada estudo.

Parte desses estudos usa dados internos das instituições, em nível de estudante. Kantorski *et al.* (2023), por exemplo, compararam métodos de aprendizado de máquina para prever a evasão em cursos presenciais de graduação, com dados de 106 cursos ao longo de três semestres. Nierotka, Salata e Martins (2023) acompanharam por seis anos 1.391 ingressantes de uma universidade federal e associaram maior risco de evasão a notas acumuladas baixas, ao ingresso fora da primeira opção de curso, ao sexo masculino e à idade mais alta no ingresso. Esse tipo de dado permite estimar o risco de cada aluno, mas só está disponível dentro da própria instituição.

Com os microdados públicos do Censo da Educação Superior, Saccaro, França e Jacinto (2019) aplicaram análise de sobrevivência a 237.697 estudantes que ingressaram em 2009 em bacharelados presenciais das áreas de Ciências, Matemática e Computação e de Engenharia, Produção e Construção, acompanhados até 2014. Os autores encontraram evasão maior em

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

instituições privadas e em cursos noturnos e menor entre estudantes com apoio financeiro, como ProUni e FIES, ou que participavam de atividades acadêmicas como estágio, monitoria e pesquisa. Vários desses fatores também estão disponíveis, de forma agregada, no arquivo de cursos do Censo e fazem parte dos atributos usados neste trabalho.

Assim, os dados públicos do Censo da Educação Superior, combinados com técnicas de *Machine Learning*, podem ser usados para gerar estimativas e indicadores que apoiem os gestores educacionais na tomada de decisão.

3. METODOLOGIA

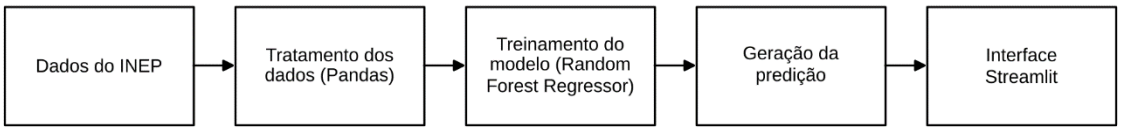
3.1. Tecnologias utilizadas

O desenvolvimento da solução empregou a linguagem Python, com as bibliotecas Pandas e NumPy para manipulação e tratamento dos dados, Scikit-Learn para a implementação do algoritmo *Random Forest Regressor*, Matplotlib para a geração dos gráficos de análise e Streamlit para o desenvolvimento da interface gráfica da aplicação.

3.2. Arquitetura da solução

A arquitetura do sistema é composta por quatro módulos principais: carregamento dos dados, pré-processamento e tratamento das informações, treinamento do modelo preditivo e apresentação dos resultados ao usuário. Inicialmente, os dados são importados para o ambiente Python, onde são submetidos às etapas de limpeza e transformação. Em seguida, as variáveis selecionadas são utilizadas no treinamento do algoritmo *Random Forest Regressor*. Por fim, os resultados são disponibilizados por meio de uma interface gráfica desenvolvida com Streamlit, conforme ilustrado na Figura 1.

Figura 1. Arquitetura da solução



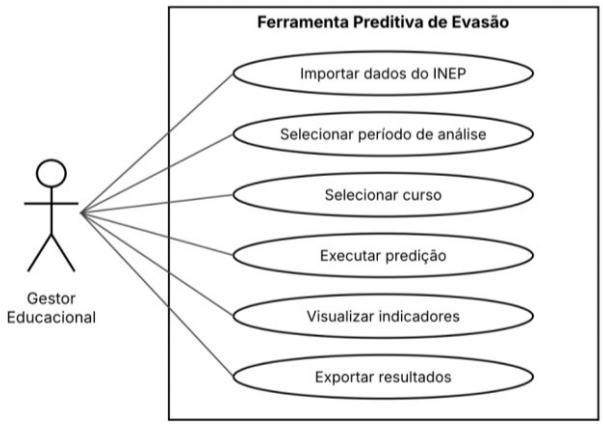
Fonte: Elaborado pelos autores, 2026.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

3.3. Diagrama de casos de uso

O sistema possui como principal ator o gestor educacional, responsável por realizar a importação dos dados, selecionar o período e o curso de interesse, executar a predição, visualizar os indicadores gerados e exportar os resultados, conforme apresentado na Figura 2.

Figura 2. Diagrama de casos de uso

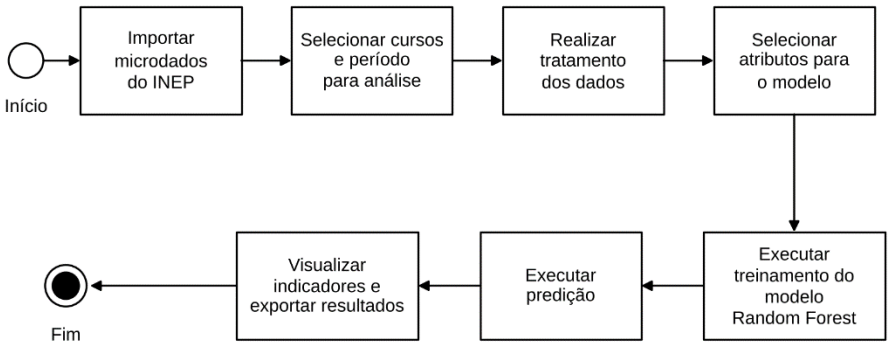


Fonte: Elaborado pelos autores, 2026.

3.4. Diagrama de atividades

A Figura 3 apresenta o fluxo de atividades da ferramenta, desde a importação dos microdados até a apresentação dos resultados ao usuário.

Figura 3. Diagrama de atividades



Fonte: Elaborado pelos autores, 2026.

3.5. Preparação dos dados e engenharia de atributos

A preparação dos dados seguiu as etapas sintetizadas no Quadro 1, implementadas nos módulos de carregamento e de pré-processamento da ferramenta.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Quadro 1. Etapas de preparação dos dados

Etapa	Descrição
Coleta	Obtenção dos microdados do INEP (2022 a 2024) e leitura em blocos, apenas com as colunas necessárias
Seleção	Aplicação do recorte (SP, bacharelado, presencial) e escolha das variáveis relevantes
Limpeza	Conversão de tipos, tratamento de valores ausentes e exclusão de cursos com menos de 30 vinculados
Transformação	Cálculo da taxa de evasão, construção de atributos relativos e montagem do painel $t \rightarrow t + 1$
Treinamento	Ajuste dos modelos com os pares 2022→2023 e busca de hiperparâmetros
Avaliação	Medição do desempenho nos pares 2023→2024

Fonte: Elaborado pelos autores, 2026.

Como os cursos têm tamanhos muito diferentes, de 30 a vários milhares de alunos vinculados, as contagens absolutas do Censo foram convertidas em proporções para que os cursos pudessem ser comparados entre si. Os atributos resultantes estão descritos no Quadro 2; a área do conhecimento segue a Classificação Internacional Normalizada da Educação (CINE), adotada pelo INEP nos microdados. Valores ausentes nos atributos numéricos foram imputados pela mediana da base de treino, e os atributos categóricos foram convertidos em variáveis binárias (*one-hot encoding*).

Quadro 2. Atributos utilizados pelo modelo

Grupo	Atributos (medidos no ano t)
Fluxo e situação acadêmica	Taxa de evasão do ano corrente; taxas de trancamento, de transferência interna, de conclusão e de ingresso (em relação ao total de vinculados)
Porte e demanda	Logaritmo do número de vinculados; candidatos por vaga; ocupação das vagas (ingressantes por vaga)
Perfil discente	Percentuais de matrículas no turno noturno, do sexo feminino, de 18 a 24 anos, com 30 anos ou mais, de alunos pretos e pardos e de egressos de escola pública
Financiamento e apoio	Percentuais de alunos com financiamento estudantil, FIES, ProUni, ingresso por reserva de vagas, apoio social e atividade extracurricular
Institucional	Categoria administrativa; organização acadêmica; área geral do conhecimento (CINE); curso na capital; curso gratuito

Fonte: Elaborado pelos autores, 2026.

3.6. Estrutura dos dados

Os dados utilizados neste trabalho foram obtidos a partir dos microdados do Censo da Educação Superior, disponibilizados anualmente pelo INEP, nas edições de 2022, 2023 e 2024 (BRASIL, 2025). Foi utilizado o arquivo de cadastro de cursos, que contém, para cada curso e ano, informações sobre vagas, inscritos, ingressantes, matrículas, concluintes, situação de

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

vínculo dos alunos, perfil dos estudantes e características da instituição, além do arquivo de cadastro de IES, empregado apenas para identificar o nome das instituições na interface.

Considerando a dimensão dessas bases, com cerca de 180 variáveis e entre 330 e 450 megabytes por ano, foi definido um recorte de estudo: cursos de bacharelado (TP_GRAU_ACADEMICO = 1), na modalidade presencial (TP_MODALIDADE_ENSINO = 1), ofertados no estado de São Paulo. A restrição à modalidade presencial decorre da estrutura dos microdados: os cursos a distância são publicados desagregados por município de polo, de modo que um mesmo curso aparece em centenas de linhas. Além disso, foram mantidos apenas cursos com pelo menos 30 alunos vinculados, pois, em turmas muito pequenas, poucas desistências produzem taxas extremas e instáveis.

Por utilizar informações de domínio público, agregadas por curso e sem possibilidade de identificação individual, a pesquisa dispensa registro e avaliação pelo sistema CEP/CONEP, conforme o art. 1º, parágrafo único, incisos III e V, da Resolução CNS nº 510/2016 (BRASIL, 2016).

Um desafio metodológico central deste trabalho consiste na operacionalização do conceito de evasão a partir dos microdados do Censo da Educação Superior, uma vez que a base é de natureza transversal, sendo coletada anualmente por curso e instituição. Conforme discutido por Coimbra, Silva e Costa (2021), a definição de evasão carece de consenso na literatura, variando conforme o recorte temporal e institucional adotado. Para a operacionalização do indicador utilizado neste estudo, adotou-se a variável QT_SIT_DESVINCULADO, disponibilizada pelos microdados do Censo da Educação Superior, que corresponde à quantidade de alunos com situação de vínculo ao curso classificada como desvinculado, distinguindo-se das demais situações possíveis de trancamento, transferência, falecimento e conclusão (BRASIL, 2025). Com base nessa variável, a taxa de evasão anual de cada curso foi calculada diretamente a partir de cada edição do Censo, sem a necessidade de acompanhar coortes ao longo de múltiplos anos, e utilizada como variável-alvo do modelo preditivo.

No Censo, a variável QT_MAT corresponde aos alunos com situação de vínculo “cursando” ou “formado” (BRASIL, 2025); assim, o denominador adotado representa o conjunto de todos os alunos que mantiveram algum vínculo com o curso no ano, conforme a Equação 1, na qual Desvinculados, Matriculados, Trancados, Transferidos e Falecidos correspondem, respectivamente, às variáveis QT_SIT_DESVINCULADO, QT_MAT, QT_SIT_TRANCADA, QT_SIT_TRANSFERIDO e QT_SIT_FALECIDO.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

$$\text{Taxa de evasão} = \frac{\text{Desvinculados}}{\text{Matriculados} + \text{Trancados} + \text{Desvinculados} + \text{Transferidos} + \text{Falecidos}} \quad (1)$$

Foi necessário, ainda, definir como utilizar essa taxa para que a ferramenta fosse efetivamente preditiva. Se o modelo estimasse a taxa de evasão de um ano com variáveis do mesmo ano, atributos como as quantidades de matriculados, trancados e transferidos integrariam o próprio cálculo da variável-alvo. Isso caracteriza vazamento de dados (KAUFMAN *et al.*, 2012) e produz métricas otimistas, sem utilidade prática, já que a taxa daquele ano seria conhecida no momento da análise. Por essa razão, o problema foi formulado de forma temporal: as características de cada curso no ano t são utilizadas para estimar a taxa de evasão do mesmo curso no ano t + 1. Os cursos foram vinculados entre as edições do Censo pelo código do curso (CO_CURSO), formando um painel de pares curso/ano.

Após o pré-processamento, detalhado na seção 3.5, foi construída uma base analítica com 13.192 registros de curso/ano (4.507 em 2022, 4.394 em 2023 e 4.291 em 2024), da qual se obteve um painel de 8.379 pares de ano corrente e ano seguinte: 4.223 pares 2022→2023 e 4.156 pares 2023→2024.

3.7. Modelagem e estratégia de validação

A validação adotou uma divisão temporal, coerente com a formulação descrita na seção 3.6: os pares 2022→2023 (4.223 cursos) compuseram a base de treino, e os pares 2023→2024 (4.156 cursos), a base de teste, de modo que o desempenho foi medido em um ano não utilizado no treinamento.

Os hiperparâmetros do *Random Forest Regressor* foram definidos por busca em grade com validação cruzada de cinco partições, realizada exclusivamente sobre a base de treino. As partições foram agrupadas por instituição (*GroupKFold*), de modo que cursos de uma mesma IES não aparecessem simultaneamente no treino e na validação, evitando que o modelo fosse favorecido por características institucionais compartilhadas. Foram avaliadas combinações de profundidade máxima das árvores (sem limite, 10 e 20), número mínimo de amostras por folha (1, 5 e 10) e fração de atributos considerada em cada divisão (0,33 e 0,6), com 300 árvores e o MAE como critério de seleção. A melhor configuração utilizou profundidade máxima 10, mínimo de 10 amostras por folha e 60% dos atributos por divisão, com MAE de validação cruzada de 6,34 pontos percentuais.

Para verificar se o modelo agrega valor em relação a alternativas mais simples, seu desempenho foi comparado ao de três referências: (i) a média, que atribui a todos os cursos a taxa média observada no treino; (ii) a persistência, que assume que a taxa do ano seguinte será igual à do ano corrente; e (iii) uma regressão linear múltipla com os mesmos atributos. Também

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

foi treinada uma variante do *Random Forest* sem a taxa de evasão do ano corrente, com o objetivo de medir quanto os demais atributos explicam isoladamente. Todos os modelos foram implementados com a biblioteca Scikit-Learn (PEDREGOSA *et al.*, 2011).

Após a avaliação, o modelo final foi treinado com todos os 8.379 pares disponíveis e aplicado aos dados de 2024, gerando a estimativa da taxa de evasão de cada curso para 2025. Para facilitar a leitura pelo gestor, as estimativas foram classificadas em três faixas de risco: baixo (inferior a 10%), moderado (de 10% a 20%) e alto (igual ou superior a 20%).

3.8. Interface da ferramenta

A interface, desenvolvida com Streamlit e executada localmente no navegador, organiza-se em cinco telas, correspondentes aos casos de uso da Figura 2: Visão geral, com indicadores agregados e a evolução da evasão por ano, área e categoria administrativa; Consulta por curso, com histórico, taxa prevista e faixa de risco do curso selecionado; Ranking e exportação, com filtros por área, categoria, município e faixa de risco e exportação da seleção em CSV; Desempenho do modelo, com as métricas de validação e a importância dos atributos; e Importar/atualizar dados, que permite indicar a pasta com novos microdados e reprocessar a base e o modelo sem conhecimento de programação. A Figura 4 apresenta a tela inicial da ferramenta. O processamento completo dos três anos de microdados, incluindo a busca de hiperparâmetros, foi executado em aproximadamente quatro minutos em um ambiente com dois núcleos de processamento.

Figura 4. Tela de visão geral da ferramenta



Fonte: Elaborado pelos autores, 2026.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

4. RESULTADOS E DISCUSSÃO

4.1. Caracterização da evasão no recorte

A Tabela 1 apresenta a síntese do indicador de evasão nos três anos analisados. A taxa agregada, que é a razão entre o total de desvinculados e o total de vinculados, variou entre 14,8% e 16,6%. Já a média das taxas por curso ficou sempre acima, entre 16,5% e 18,5%, o que indica que os cursos menores tendem a ter evasão mais alta que os maiores.

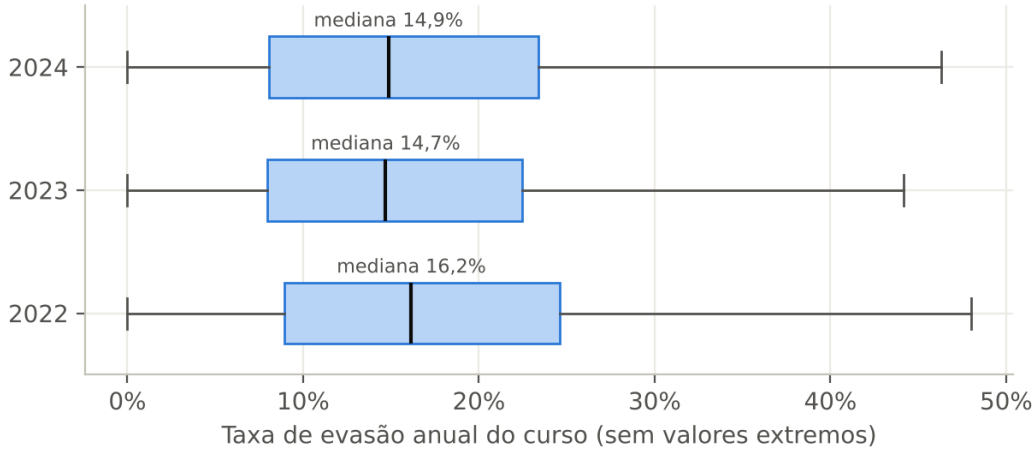
Tabela 1. Indicadores de evasão por ano (bacharelados presenciais de SP com 30 ou mais vinculados)

Ano	Cursos	Vinculados	Desvinculados	Taxa agregada	Média por curso	Mediana por curso
2022	4.507	1.303.215	216.589	16,6%	18,5%	16,2%
2023	4.394	1.243.911	183.598	14,8%	16,5%	14,7%
2024	4.291	1.217.214	187.130	15,4%	17,3%	14,9%

Fonte: Elaborado pelos autores com base em dados do INEP (BRASIL, 2025).

A Figura 5 mostra que a distribuição das taxas por curso é assimétrica: metade dos cursos apresenta evasão anual entre aproximadamente 8% e 24%, mas há uma cauda de cursos com taxas muito superiores, incluindo casos próximos de 100%. Esses casos extremos correspondem, em sua maioria, a cursos sem novos ingressantes e com praticamente todos os alunos desvinculados, situação típica de cursos em processo de desativação.

Figura 5. Distribuição da taxa de evasão anual por curso

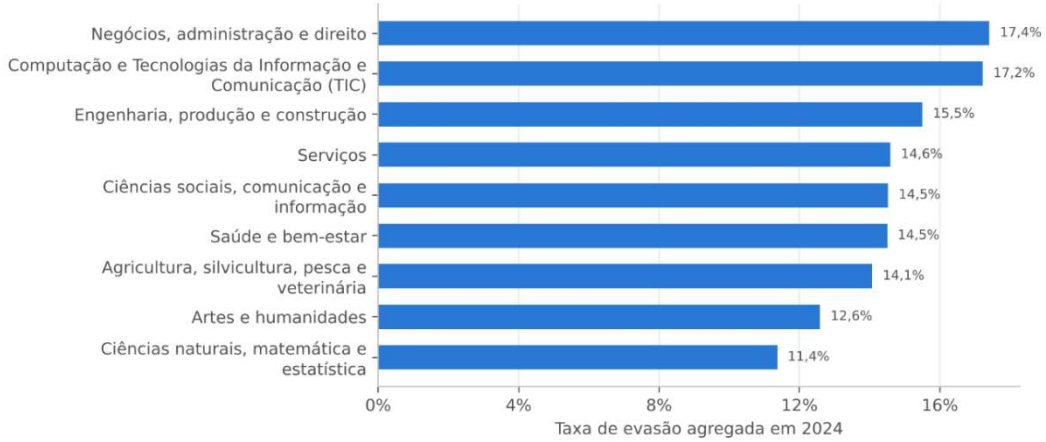


Fonte: Elaborado pelos autores com base em dados do INEP (BRASIL, 2025).

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Também se observam diferenças expressivas entre as categorias administrativas. Em 2024, a taxa agregada foi de 6,7% nas instituições públicas, 11,5% nas privadas sem fins lucrativos e 21,3% nas privadas com fins lucrativos, padrão que se repetiu nos três anos analisados. Entre as áreas do conhecimento (Figura 6), as maiores taxas ocorreram em Negócios, administração e direito (17,4%) e em Computação e Tecnologias da Informação e Comunicação (17,2%), e as menores em Ciências naturais, matemática e estatística (11,4%) e em Artes e humanidades (12,6%). Ainda assim, as diferenças entre as áreas são menores que as observadas entre categorias administrativas, o que também se reflete na importância dos atributos (seção 4.4).

Figura 6. Taxa de evasão agregada por área do conhecimento (CINE), 2024



Fonte: Elaborado pelos autores com base em dados do INEP (BRASIL, 2025).

4.2. Desempenho dos modelos

A Tabela 2 apresenta o desempenho dos modelos na base de teste (pares 2023→2024).

Tabela 2. Desempenho dos modelos na base de teste

Modelo	MAE (p.p.)	MSE	RMSE (p.p.)	R^2
Média (referência)	9,52	0,0175	13,24	−0,003
Persistência (taxa atual)	7,24	0,0168	12,97	0,037
Regressão Linear	6,84	0,0132	11,51	0,242
Random Forest	6,53	0,0131	11,46	0,248
Random Forest sem taxa atual	6,79	0,0128	11,32	0,267

Fonte: Elaborado pelos autores, 2026.

O *Random Forest* obteve o menor erro médio absoluto, 6,53 pontos percentuais (p.p.), o que representa redução de 31,5% em relação à previsão pela média (9,52 p.p.), de 9,9% em relação à persistência (7,24 p.p.) e de 4,6% em relação à regressão linear (6,84 p.p.). O R^2 de

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

0,248 indica que o modelo explica cerca de um quarto da variância da taxa de evasão do ano seguinte.

O resultado do modelo de persistência ajuda a entender as métricas. Repetir a taxa do ano anterior reduz substancialmente o MAE em relação à média, porém o R^2 permanece próximo de zero (0,037). Esse comportamento decorre dos erros elevados do modelo nos cursos em que a evasão varia bruscamente de um ano para o outro, os quais pesam mais nas métricas quadráticas. Na variante do *Random Forest* sem a taxa corrente observou-se o oposto: o MAE foi maior (6,79 p.p.) e o R^2 um pouco mais alto (0,267). Uma explicação plausível é que os atributos estruturais do curso (perfil dos alunos, financiamento e características da instituição) ajudam principalmente a reduzir os erros grandes, enquanto a taxa do ano corrente melhora a previsão nos casos mais comuns. Como o MAE foi o critério escolhido para selecionar o modelo e é a métrica mais fácil de interpretar para o gestor, a ferramenta usa o *Random Forest* completo.

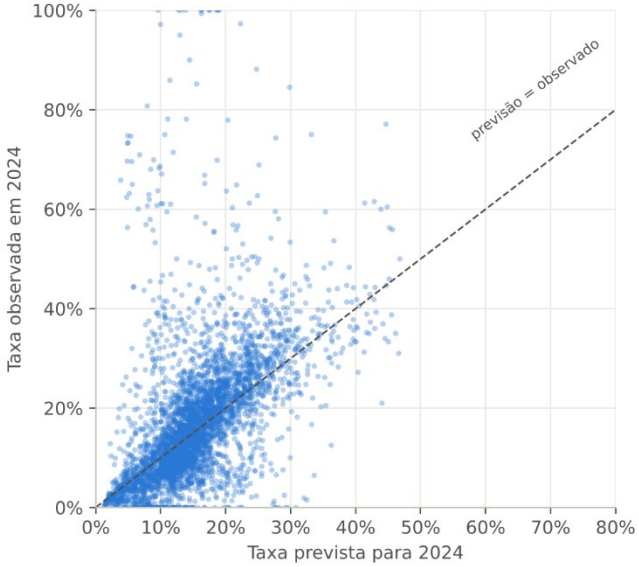
O R^2 obtido deve ser interpretado à luz do tipo de problema. A taxa de evasão de um curso em um ano depende de fatores que não estão nos microdados, como mudanças na gestão, políticas de retenção, a situação econômica dos alunos e mudanças no currículo, além de variações aleatórias nos cursos menores. Por essa razão, o modelo constitui uma estimativa de referência, baseada exclusivamente em dados públicos, e deve ser utilizado em conjunto com o acompanhamento realizado pela própria instituição.

4.3. Análise dos erros e classificação de risco

A Figura 7 compara as taxas previstas e observadas para os 4.156 cursos da base de teste. A mediana do erro absoluto foi de 3,9 p.p.; 59,8% dos cursos tiveram erro inferior a 5 p.p. e 83,1%, inferior a 10 p.p. Os maiores erros concentram-se em cursos cuja evasão observada foi muito superior à prevista (pontos acima da diagonal), incluindo 16 cursos com taxa igual ou superior a 90% em 2024, majoritariamente os cursos em desativação discutidos na seção 4.1. O modelo também tende a suavizar os valores extremos: a média das previsões (15,5%) ficou abaixo da média observada (17,4%), o que é esperado, já que o *Random Forest* calcula a média das previsões de várias árvores e por isso dificilmente gera valores muito altos ou muito baixos.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Figura 7. Taxas de evasão previstas e observadas na base de teste (Random Forest)



Fonte: Elaborado pelos autores, 2026.

Para o gestor, além do valor previsto, interessa saber se o modelo consegue apontar os cursos com evasão alta. A Tabela 3 cruza as faixas de risco previstas com as observadas.

Tabela 3. Faixas de risco previstas e observadas na base de teste

Prevista \ Observada	Baixo (< 10%)	Moderado (10–20%)	Alto (≥ 20%)	Total
Baixo	641	123	86	850
Moderado	603	1.162	618	2.383
Alto	77	112	734	923
Total	1.321	1.397	1.438	4.156

Fonte: Elaborado pelos autores, 2026.

A faixa prevista coincidiu com a observada em 61,0% dos cursos. Dos 923 cursos classificados como de risco alto, 734 (79,5%) registraram evasão igual ou superior a 20%, e apenas 77 (8,3%) ficaram abaixo de 10%. Em contrapartida, o modelo identificou 51,0% dos cursos que tiveram evasão alta e classificou grande parte dos demais como risco moderado. Assim, quando o modelo indica risco alto, a indicação costuma ser correta, mas nem todos os cursos com evasão alta são captados nessa faixa.

O erro também variou entre categorias administrativas: o MAE foi de 3,5 p.p. nas instituições públicas, 6,7 p.p. nas privadas sem fins lucrativos e 7,4 p.p. nas privadas com fins lucrativos. Neste último grupo, que concentra quase metade dos cursos do teste (1.990) e as maiores taxas de evasão, o modelo subestimou a média observada (19,1% previstos contra

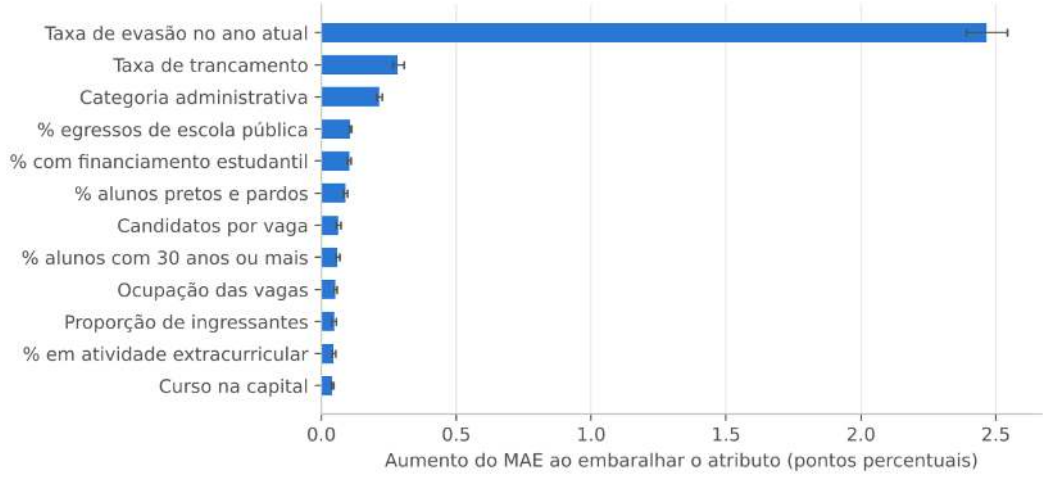
Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

23,3% observados), o que é compatível com uma maior volatilidade da evasão entre anos nessas instituições.

4.4. Importância dos atributos

A Figura 8 apresenta a importância dos atributos calculada por permutação na base de teste. A taxa de evasão do ano corrente foi o atributo mais importante, com ampla vantagem sobre os demais: embaralhá-la aumenta o MAE em 2,47 p.p., o que confirma a forte persistência da evasão ao longo do tempo. Em seguida aparecem a taxa de trancamento (0,28 p.p.), que pode indicar alunos em vias de abandonar o curso, e a categoria administrativa da instituição (0,22 p.p.), coerente com as diferenças observadas na seção 4.1. Depois vêm atributos do perfil socioeconômico dos alunos, como os percentuais de egressos de escola pública, de alunos com financiamento estudantil e de alunos pretos e pardos, com contribuições menores. Esse padrão se aproxima do encontrado por Saccaro, França e Jacinto (2019), que também observaram mais evasão em instituições privadas e associação entre apoio financeiro e permanência no curso.

Figura 8. Importância dos atributos por permutação (doze mais relevantes)



Fonte: Elaborado pelos autores, 2026.

Esses resultados indicam associações úteis para a previsão e não permitem conclusões sobre causa e efeito. O peso do percentual de egressos de escola pública, por exemplo, pode estar ligado a dificuldades socioeconômicas dos alunos ou a características das instituições que recebem esse público. A literatura também trata a evasão como um fenômeno com muitas causas e sem definição única (COIMBRA; SILVA; COSTA, 2021).

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

4.5. Aplicação: estimativa para 2025

O modelo final, treinado com todos os pares disponíveis, foi aplicado aos 4.291 cursos do recorte em 2024. A taxa média estimada para 2025 foi de 17,0%, com mediana de 15,9%. Na classificação de risco, 749 cursos (17,5%) foram enquadrados na faixa baixa, 2.162 (50,4%) na moderada e 1.380 (32,2%) na alta. Na ferramenta, essas estimativas podem ser filtradas por área, categoria administrativa, município e faixa de risco e exportadas, permitindo ao gestor priorizar cursos para ações de acompanhamento e de permanência estudantil.

4.6. Limitações

Algumas limitações devem ser consideradas. Primeiro, a unidade de análise é o curso, e não o estudante: a ferramenta estima a evasão esperada de um curso, mas não identifica quais alunos estão em risco, o que exigiria dados internos da instituição em nível de estudante, como os usados por Nierotka, Salata e Martins (2023). Segundo, a disponibilidade de apenas três edições do Censo permitiu um único par de anos para treino e outro para teste; com mais edições, seria possível treinar com séries mais longas e verificar a estabilidade do desempenho ao longo do tempo. Terceiro, o indicador adotado mede a evasão anual do curso, que não equivale à evasão de uma coorte ao longo de toda a graduação, e a situação de desvinculado pode incluir saídas para cursos de outras instituições, que a literatura trata como mobilidade e não necessariamente como abandono do sistema. Por fim, o recorte restrito a cursos presenciais de bacharelado em São Paulo limita a generalização dos resultados, embora a ferramenta permita reprocessar os dados para outras unidades da federação. Além disso, o poder explicativo moderado do modelo (R^2 de 0,248) recomenda cautela quanto à generalização das previsões, que devem ser interpretadas como estimativas de referência e não como valores exatos.

5. CONSIDERAÇÕES FINAIS

A decisão que mais influenciou os resultados deste trabalho foi a forma de formular o problema. Na proposta inicial, a taxa de evasão de um ano seria estimada com variáveis do mesmo ano, o que produziria métricas altas sem utilidade, porque essa taxa já seria conhecida no momento da análise. Ao prever o ano seguinte e testar o modelo em um ano que ele não tinha visto, os resultados foram mais modestos, porém passaram a representar o desempenho que a ferramenta efetivamente alcança em uso real.

Com essa formulação, a resposta ao problema de pesquisa é positiva, com ressalvas. É possível construir uma ferramenta preditiva apenas com os microdados do Censo, e ela

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

apresenta erro menor que o das alternativas simples: o erro médio caiu de 9,5 para 6,5 pontos percentuais em relação à previsão pela média e ficou abaixo do obtido ao repetir a taxa do ano anterior. A hipótese foi confirmada em parte. As características dos cursos ajudam a prever a evasão, mas a maior parte da variação de um ano para o outro continua sem explicação (R^2 de 0,248). Isso mostra um limite dos dados públicos: o Censo registra quem saiu do curso, mas não registra por que saiu, nem o que a instituição fez para evitar a saída.

O resultado mais útil para a gestão foi a indicação de risco alto, que acertou em quase oito de cada dez cursos sinalizados. Esse uso parece mais adequado do que tratar a taxa prevista como um número exato: a ferramenta serve para indicar quais cursos merecem atenção prioritária, e a investigação das causas continua dependendo do conhecimento que cada instituição tem dos seus cursos e alunos. Recomenda-se, em estudos futuros, ampliar o número de edições do Censo e testar a ferramenta em outras unidades da federação e modalidades de ensino, a fim de verificar a generalização dos resultados.

DECLARAÇÃO DE USO DE INTELIGÊNCIA ARTIFICIAL

Este trabalho utilizou a ferramenta de inteligência artificial generativa Claude (Anthropic) como apoio no desenvolvimento do código-fonte da ferramenta, no processamento e na análise dos dados, na elaboração das figuras e na redação, revisão e tradução do texto. Os autores definiram o tema, os objetivos e o recorte da pesquisa, revisaram todo o conteúdo e se responsabilizam integralmente por ele.

CONFLITO DE INTERESSES E FINANCIAMENTO

Os autores declaram não haver conflitos de interesses. A pesquisa não recebeu financiamento.

REFERÊNCIAS

BRASIL. Conselho Nacional de Saúde. Resolução nº 510, de 7 de abril de 2016. Dispõe sobre as normas aplicáveis a pesquisas em Ciências Humanas e Sociais. **Diário Oficial da União**: seção 1, Brasília, DF, p. 44-46, 24 maio 2016.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). **Censo da Educação Superior**: microdados. Brasília, DF: INEP, 2025. Disponível em: https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/censo-da-educacao-superior. Acesso em: 30 maio 2026.

BREIMAN, Leo. Random forests. **Machine Learning**, v. 45, n. 1, p. 5-32, 2001. DOI: https://doi.org/10.1023/A:1010933404324.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

COIMBRA, Camila Lima; SILVA, Leonardo Barbosa e; COSTA, Natália Cristina Dreossi. A evasão na educação superior: definições e trajetórias. **Educação e Pesquisa**, São Paulo, v. 47, e228764, 2021. DOI: https://doi.org/10.1590/S1678-4634202147228764.

FACELI, Katti; LORENA, Ana Carolina; GAMA, João; CARVALHO, André Carlos Ponce de Leon Ferreira de. **Inteligência Artificial**: uma abordagem de aprendizado de máquina. Rio de Janeiro: LTC, 2011.

KANTORSKI, Gustavo Zanini; MARTINS, Ricardo Zimmermann; BALEJO, Arthur; FRICK, Marcio. Mineração de dados educacionais para predição da evasão em cursos de graduação presenciais no ensino superior. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO (SBIE), 34., 2023, Passo Fundo. **Anais** [...]. Porto Alegre: Sociedade Brasileira de Computação, 2023. p. 1133-1142. DOI: https://doi.org/10.5753/sbie.2023.233746.

KAUFMAN, Shachar; ROSSET, Saharon; PERLICH, Claudia; STITELMAN, Ori. Leakage in data mining: formulation, detection, and avoidance. **ACM Transactions on Knowledge Discovery from Data**, v. 6, n. 4, art. 15, 2012. DOI: https://doi.org/10.1145/2382577.2382579.

KUHN, Max; JOHNSON, Kjell. **Applied Predictive Modeling**. New York: Springer, 2013. DOI: https://doi.org/10.1007/978-1-4614-6849-3.

MACIEL, Carina Elisabeth; CUNHA JÚNIOR, Mauro; LIMA, Tatiane da Silva. A produção científica sobre permanência e evasão na educação superior no Brasil. **Educação e Pesquisa**, São Paulo, v. 45, e198669, 2019. DOI: https://doi.org/10.1590/S1678-4634201945198669.

NIEROTKA, Rosileia Lucia; SALATA, Andre; MARTINS, Melina Klitzke. Fatores associados à evasão no ensino superior: um estudo longitudinal. **Cadernos de Pesquisa**, São Paulo, v. 53, e09961, 2023. DOI: https://doi.org/10.1590/198053149961.

PEDREGOSA, Fabian *et al.* Scikit-learn: machine learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825-2830, 2011.

SACCARO, Alice; FRANÇA, Marco Túlio Aniceto; JACINTO, Paulo de Andrade. Fatores associados à evasão no ensino superior brasileiro: um estudo de análise de sobrevivência para os cursos das áreas de Ciência, Matemática e Computação e de Engenharia, Produção e Construção em instituições públicas e privadas. **Estudos Econômicos**, São Paulo, v. 49, n. 2, p. 337-373, 2019. DOI: https://doi.org/10.1590/0101-41614925amp.

SILVESTRI, Felipe; SOUZA, Vanessa Faria de; VIEIRA, Andrws. Mineração de dados educacionais: uma análise sobre os preditores da evasão no ensino superior. **Revista Educar Mais**, v. 9, p. 1-23, 2025. DOI: https://doi.org/10.15536/reducarmais.9.2025.4120.

STARK, Julia Wickboldt; PINTO, Rodrigo Serpa. Evasão no ensino superior: uma revisão sistemática das abordagens metodológicas e teóricas no Brasil. **Regae: Revista de Gestão e Avaliação Educacional**, Santa Maria, v. 14, n. 23, e90900, 2025. DOI: https://doi.org/10.5902/2318133890900.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.