

Aplicação para previsão de abandono estudantil

Alex Se Yoon Koo¹, Gabriel Novaes², João Victor Pacheco³

¹Faculdade de Computação e Informática - Universidade Presbiteriana Mackenzie (UPM)

Abstract. *Student dropout is a complex and multifaceted phenomenon, which refers to students giving up or interrupting their studies before completing a course. The causes of this dropout can vary, including socioeconomic factors, mental health problems, lack of motivation, lack of educational support, challenges in balancing work and study, among others. This problem affects both basic and higher education, and has serious implications for individual and social development. Student dropout can result in economic losses, lower professional qualifications, increased social inequalities and limitations in career opportunities. Therefore, this research aims to create a classification algorithm to define whether a student will complete the course or not.*

Resumo. *O abandono estudantil é um fenômeno complexo e multifacetado, que se refere à desistência ou interrupção dos estudos por parte dos alunos antes da conclusão de um curso. As causas desse abandono podem variar, incluindo fatores socioeconômicos, problemas de saúde mental, desmotivação, falta de suporte educacional, desafios na conciliação de trabalho e estudo, entre outros. Esse problema afeta tanto o ensino básico quanto o superior, e tem implicações graves para o desenvolvimento individual e social. O abandono estudantil pode resultar em perdas econômicas, menor qualificação profissional, aumento das desigualdades sociais e limitações nas oportunidades de carreira. Com isso, esta pesquisa tem como objetivo de criar um algoritmo de classificação para definir se um aluno concluirá o curso ou não.*

1. Introdução

1.1. Contextualização

No cotidiano que vivemos a educação apresenta ser algo importante ou crucial, desenvolvendo conhecimento e pessoas, além de expandir o horizontal. Porém isso não apresenta ser algo que ocorre em todo o momento uma vez que mesmo em situações de educação ocorre casos de evasão, onde perdemos talentos por motivos econômicos ou sociais.

1.2. Justificativa

Diante disso, com a evasão escolar aumentando cada vez mais com o tempo, alunos deixando os estudos, perdemos a uma quantidade de talentos potenciais para a pesquisa ou acadêmico como também no âmbito profissional, fazendo-se assim a desvalorização da cultura brasileira nesse requisito.

1.3. Objetivo

Esta pesquisa tem como objetivo de fazer uma classificação dos alunos da chance de graduar ou não o curso de bacharelado presente de acordo com os seus atributos sociais e econômicos como uma forma de prevenir a evasão escolar.

1.4. Opção do projeto/Objetivo Específico

Como objetivo específico, temos como alvo de, além da classificação, buscar entender quais são os principais fatores para evasão escolar e a correlação entre eles, e se diferentes algoritmos de classificação afeta a probabilidade final.

2. Descrição do Problema

A educação desempenha um papel fundamental no desenvolvimento de indivíduos e sociedades, contribuindo para a expansão do conhecimento e a formação de talentos. No entanto, a realidade educacional brasileira enfrenta desafios significativos, sendo a evasão escolar um dos problemas mais críticos.

Esse fenômeno ocorre quando estudantes abandonam seus estudos antes da conclusão, resultando na perda de potenciais talentos acadêmicos e profissionais. As causas são multifacetadas, abrangendo fatores econômicos, sociais e estruturais, que impactam diretamente a trajetória educacional de milhares de jovens. A evasão escolar não apenas prejudica o futuro dos estudantes, mas também contribui para a desvalorização cultural e econômica do país, restringindo oportunidades de inovação e desenvolvimento. Além disso, a ausência de políticas eficazes para prevenir e mitigar essa problemática agrava o cenário, perpetuando desigualdades e limitando o acesso à educação de qualidade.

Diante desse contexto, torna-se crucial identificar e compreender os fatores que levam à evasão escolar, como forma de desenvolver estratégias de intervenção. Neste artigo, propõe-se classificar as chances de um estudante concluir ou não um curso de bacharelado, analisando seus atributos sociais e econômicos, além de explorar a relação entre esses fatores e avaliar o impacto de diferentes algoritmos de classificação no resultado. Assim, busca-se não apenas entender as raízes do problema, mas também contribuir para a criação de soluções que promovam a permanência e o sucesso acadêmico.

3. Dataset

O dataset que utilizamos foi dados csv disponível na Kaggle sobre evasão escolar <https://www.kaggle.com/datasets/syedfaizanalii/predict-students-dropout-and-academic-success/data>. O dataset possui os seguintes atributos:

Estado Civil (Marital Status): O estado civil do estudante (por exemplo, solteiro, casado, divorciado).

Modo de Inscrição (Application Mode): Refere-se ao modo ou tipo de inscrição que o estudante submeteu para se matricular no curso.

Ordem de Inscrição (Application Order): Indica a ordem em que o estudante se candidatou ao curso. Por exemplo, se foi a primeira, segunda ou terceira opção do estudante.

Curso (Course): O curso ou programa de graduação em que o estudante está matriculado (por exemplo, Ciência da Computação, Engenharia, etc.).

Turno de Frequência (Daytime/Evening Attendance): Especifica se o estudante frequenta o curso durante o dia ou à noite, representando seu horário de frequência.

Qualificação Anterior (Previous Qualification): O tipo de qualificação acadêmica que o estudante tinha antes de se matricular no curso (por exemplo, diploma de ensino médio, curso técnico).

Qualificação Anterior (Nota) [Previous Qualification (Grade)]: A nota final ou pontuação associada à qualificação anterior do estudante.

Nacionalidade (Nationality): A nacionalidade do estudante.

Qualificação da Mãe (Mother's Qualification): O nível mais alto de qualificação acadêmica alcançado pela mãe do estudante.

Qualificação do Pai (Father's Qualification): O nível mais alto de qualificação acadêmica alcançado pelo pai do estudante.

De acordo com a análise exploratória de dados, podemos observar que metade dos alunos tem chance de terminar o curso de graduação e 0.32 de porcentagem em largar a faculdade de acordo com a figura 3.

Também é possível observar a distribuição dos dados de acordo com a figura 4, onde criamos gráficos entre dois atributos quaisquer para ver como está a distribuição dos dados.

Também realizamos um bucket do target label para cada coluna para observar o comportamento dos dados para cada coluna como mostra na figura 1 e 2.

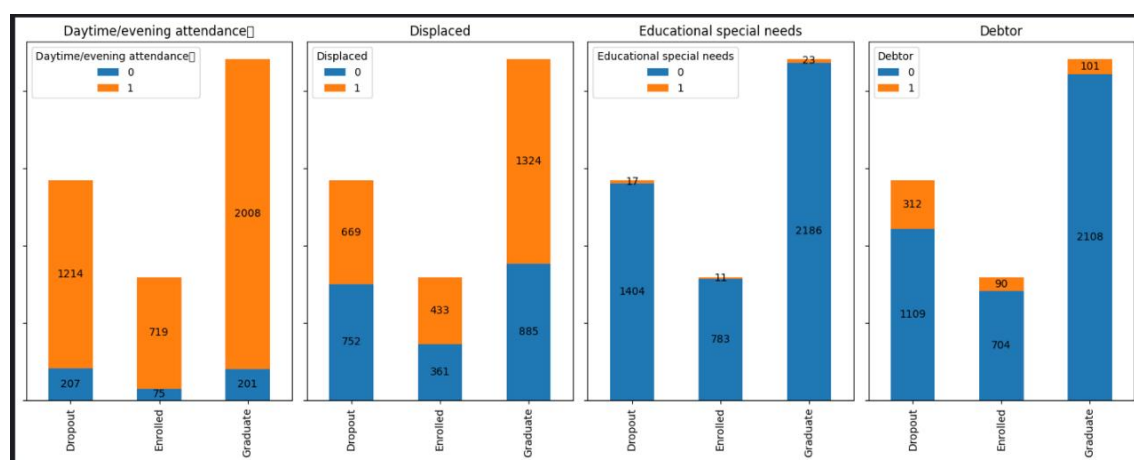


Figura 1. Bucket dos alunos de acordo com cada coluna.

4. Metodologia e Resultado Esperado

Nesta seção buscamos desdobrar a metodologia adotada e os materiais e estimar o resultado esperado da nossa pesquisa.

Como primeira etapa, buscamos estudos pesquisas anteriormente já realizadas como uma forma de referência e definir quais serão os próximos passos, em seguida na segunda etapa, buscamos coletar e interpretar os dados usando análise exploratória dos dados e definir quais os resultados que podemos obter. Na terceira etapa, realizamos o pré-processamento dos dados, limpando os dados, seja na sua duplicidade ou formatação, e consequentemente, na quarta etapa separamos os dados e realizamos a validação cruzada

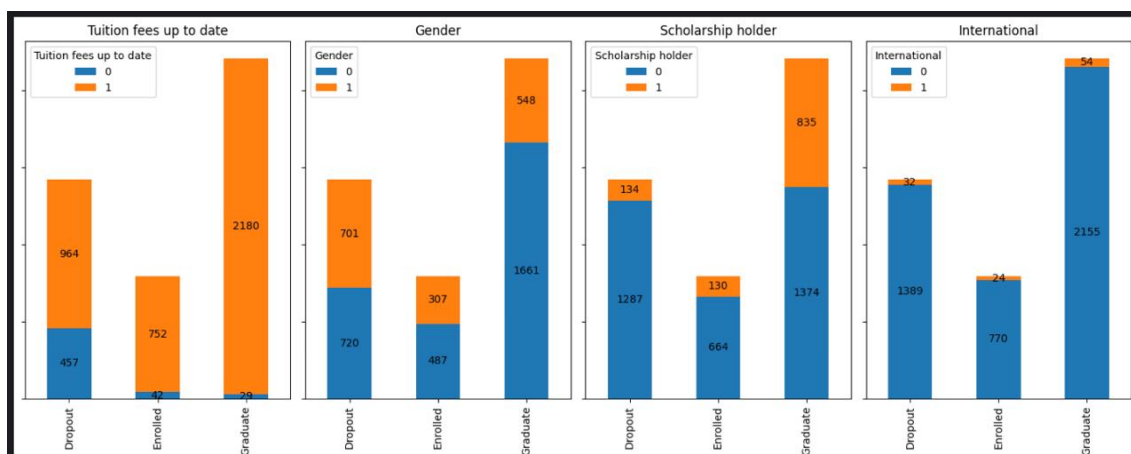


Figura 2. Bucket dos alunos de acordo com cada coluna.

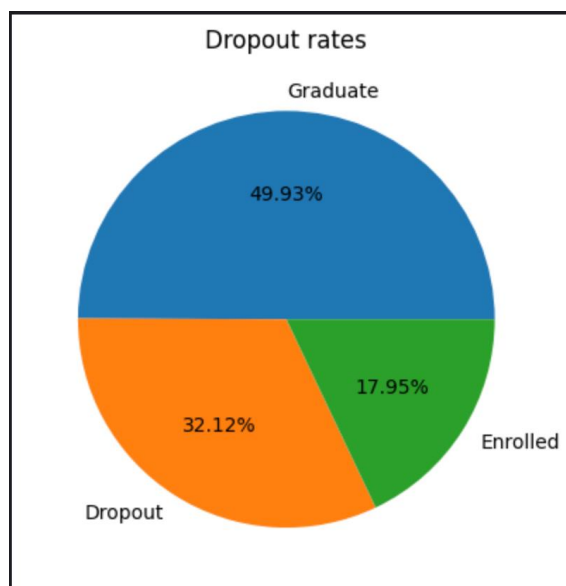


Figura 3. Gráfico de pizza mostrando a porcentagem de alunos em no label target.

como uma forma de diminuir as chances de overfitting durante o treinamento. Na quinta etapa, buscamos algoritmos que podemos utilizar e comparar entre eles como também as métricas que iremos utilizar para avaliar o nosso modelo. Na sexta etapa, buscamos implementar o nosso algoritmo para treinamento e previsão e como última etapa analisar os dados.

5. Resultados

A aplicação do algoritmo de classificação proposta neste estudo apresentou resultados relevantes para a análise da evasão estudantil. Através da exploração dos dados disponíveis, foi possível identificar padrões e correlações entre variáveis sociais e econômicas que influenciam na conclusão ou desistência do curso pelos estudantes. A validação cruzada dos modelos demonstrou eficácia na previsão de abandono escolar, reduzindo o risco de overfitting durante o treinamento. A análise do desempenho dos algoritmos indicou que



Figura 4. Correlação entre duas colunas ou atributos.

as métricas utilizadas, como precisão, recall e F1-score, foram adequadas para avaliar a qualidade das predições realizadas (Figura 5).

Além disso, a distribuição e a visualização gráfica dos dados revelaram insights importantes sobre os principais fatores associados à evasão escolar. Por exemplo, variáveis de unidades curriculares mostraram-se altamente correlacionadas com a probabilidade de um estudante concluir o curso (Figura 6).

6. Considerações Finais

Este estudo destacou a relevância de compreender os fatores que influenciam a evasão escolar e a importância de adotar estratégias baseadas em dados para mitigar esse problema. A utilização do algoritmo de classificação Random Tree permitiu identificar variáveis socioeconômicas e educacionais que impactam diretamente na permanência dos estudantes nos cursos em questão. Os resultados obtidos demonstram que a análise preditiva é uma ferramenta poderosa para antecipar casos de possível abandono, possibilitando intervenções direcionadas e mais eficazes. Conclui-se que, para reduzir a evasão escolar, é fundamental não apenas aplicar soluções tecnológicas, mas também promover políticas

```

Accuracy: 0.7638418079096045
Classification Report:

```

	precision	recall	f1-score	support
0	0.82	0.76	0.79	297
1	0.55	0.26	0.35	157
2	0.76	0.95	0.85	431
accuracy			0.76	885
macro avg	0.71	0.66	0.66	885
weighted avg	0.75	0.76	0.74	885

Figura 5. Relatório de classificacao

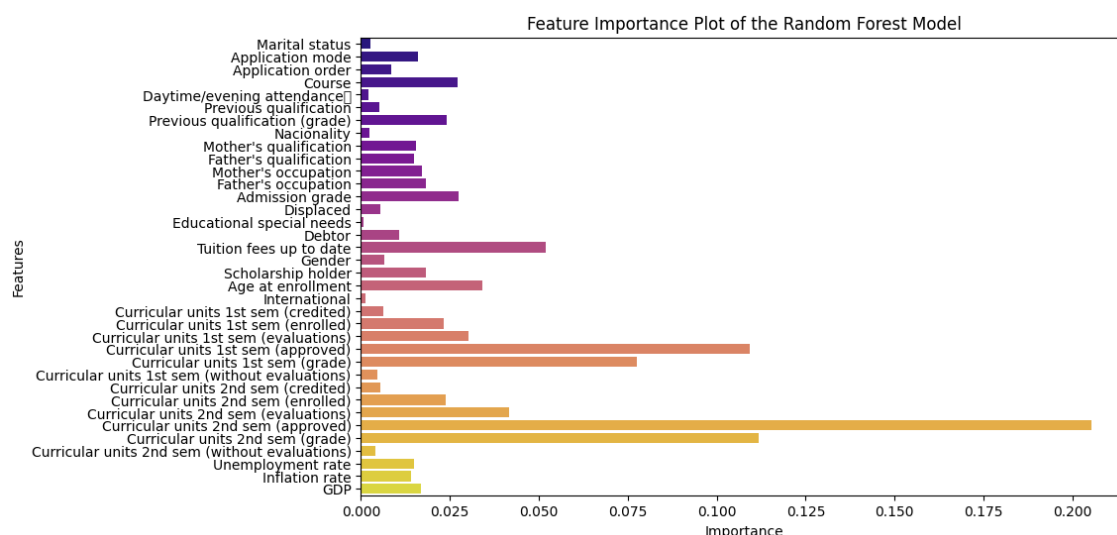


Figura 6. Relevância dos atributos

públicas e ações institucionais que garantam suporte acadêmico, emocional e financeiro aos estudantes, criando condições mais favoráveis para sua formação e sucesso profissional.

7. Referência e Bibliografia

7.1. Fatores associados ao abandono escolar no ensino médio público de Minas

Gerais

Esse artigo aponta os principais fatores que influenciam o abandono escolar no ensino médio, onde os dados se baseiam em uma série histórica constituída a partir do levantamento das pesquisas nacionais por amostra de Domicílios.

7.2. Abandono e evasão escolar sobre a otica dos sujeitos envolvidos

O artigo busca refletir sobre as concepções relativas ao abandono e à evasão escolar por parte de agentes públicos que atuam junto à população juvenil em um município pde porte médio do interior paulista.

7.3. A problemática da evasão escolar: Uma revisão bibliográfica integrativa

O artigo visa a educação no Brasil que enfrenta diversos problemas que tem onde ele aborda a importância do direito a educação e mostra as principais causas que estão ligadas a evasão escolar.

7.4. Abandono escolar: Entendendo as causas e buscando soluções

Nesse artigo ele mostra dados de acordo com o senso e a OCDE onde ele explica a realidade do abandono escolar e fala sobre a diferença entre abandono escolar e evasão escolar.

7.5. Evasão e abandono escolar na educação básica no Brasil:fatores,causas e possíveis consequências

O artigo objetiva trazer algumas considerações sobre a evasão e abandono escolar na educação brasileira e mostra diversos fatores entre eles é dizer que existem três dimensões conceituais para a investigação da evasão escolar.

Referências

Brasil: fatores, causas e possíveis consequências Dropout and abandonment of school in Brazil: factors, causes and possible consequences Raimundo Barbosa Silva Filhoa. Ronaldo Marcos de Lima Araújo: [s.n.].

RAMOS, A. C.; GONÇALVES JUNIOR, O. Abandono e evasão escolar sob a ótica dos sujeitos envolvidos. **Educação e Pesquisa**, v. 50, p. e268037, 2024.

SENNA, I. A. **Abandono escolar: entendendo as causas e buscando soluções**. Disponível em: <https://institutoayrtonsenna.org.br/abandono-escolar/>. Acesso em: 29 set. 2024.

SOARES, T. M. et al. Fatores associados ao abandono escolar no ensino médio público de Minas Gerais. **Educação e Pesquisa**, v. 41, n. 3, p. 757–772, 2015.

LINO, R. E.. A Problemática da evasão escolar: uma revisão bibliografia interativa. **PUC-Goiás**, 2020.

8. Link do Github e vídeo do YouTube

GitHub: <https://github.com/alexsky1502/projeto-ia-ivan>

YouTube: https://www.youtube.com/watch?v=s3eI7HHa5VI&ab_channel=AlexSeYoonKoo