

Aplicação de Modelos de Aprendizado de Máquina na Predição da Desistência Acadêmica em Cursos de TI do IFPE

Yanka Mirelly Ayala de Sales Silva¹

Verlaynne Kelley da Hora Rocha Araujo¹

Diogo Lopes da Silva¹

¹Instituto Federal de Pernambuco

Palmares - PE - Brasil

ymass@discente.ifpe.edu.br

{verlaynne.araujo, diogo.lopes}@palmares.ifpe.edu.br

Resumo

A evasão acadêmica constitui um dos principais desafios enfrentados pelas instituições de ensino, especialmente nos cursos de Ciências Exatas, que apresentam altos índices de desistência. Este estudo, aplicado ao contexto do IFPE, propõe a utilização de técnicas de aprendizado de máquina para prever a evasão estudantil, com base em dados acadêmicos, demográficos e socioeconômicos de alunos entre os períodos de 2001.1 e 2025.1. Após o pré-processamento dos dados, que incluiu codificação de variáveis, tratamento de valores ausentes e balanceamento das classes com SMOTE, foram treinados três modelos supervisionados: *Decision Tree*, *Extra Trees* e *Multilayer Perceptron* (MLP). Os modelos foram ajustados via Grid Search e avaliados por métricas como acurácia, precisão, *recall* e F1-score. O MLP apresentou o melhor desempenho, destacando-se pela capacidade de identificar com maior precisão os alunos propensos à evasão. Os resultados demonstram o potencial da inteligência artificial como ferramenta de apoio à gestão educacional, favorecendo a adoção de medidas preventivas mais eficazes.

Palavras-chave: Redes neurais artificiais. *Decision Tree*. *Extra Trees*. Evasão acadêmica. Aprendizado de máquina. IFPE.

Abstract

Dropout rates are one of the main challenges faced by educational institutions, especially in

the Exact Sciences, which have high dropout rates. This study, applied to the IFPE context, proposes the use of machine learning techniques to predict student dropout rates based on academic, demographic, and socioeconomic data from students between the 2001–2025 academic years. After data preprocessing, which included variable coding, missing value processing, and class balancing with SMOTE, three supervised models were trained: Decision Tree, Extra Trees, and Multilayer Perceptron (MLP). The models were adjusted using Grid Search and evaluated using metrics such as accuracy, precision, recall, and F1-score. The MLP performed best, standing out for its ability to more accurately identify students prone to dropout. The results demonstrate the potential of artificial intelligence as a tool to support educational management, favoring the adoption of more effective preventive measures.

Keywords: Artificial neural networks. Decision trees. Extra trees. Academic dropout. Machine learning. IFPE.

1. Introdução

A evasão acadêmica é um problema complexo que impacta significativamente as instituições de ensino, especialmente em cursos de exatas, estendido pela complexidade das disciplinas que resultam em desafios aos estudantes. No contexto do Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco (IFPE), o problema também é um desafio, afetando todas as modalidades de ensino (Integrado, Subsequente e Superior), com reflexos diretos no desempenho institucional e no futuro dos alunos (Garcia e Gomes, 2022).

A evasão não é apenas um indicador de insucesso acadêmico, mas também um reflexo de questões socioeconômicas, psicológicas e pedagógicas que permeiam a vida dos estudantes (Lopes, 2022). De acordo com (Salata, 2019), fatores como dificuldades financeiras, necessidade de conciliar trabalho e estudo, baixa adaptação ao curso e à instituição, além de deficiências na formação básica, estão entre as principais causas desse fenômeno.

No ensino técnico federal, como o oferecido pelo IFPE, a evasão afeta diretamente as políticas públicas de inclusão e formação profissional. Segundo (Souza e Freitas, 2021), cursos da área de exatas apresentam maiores índices de desistência, o que reforça a necessidade de ações preventivas fundamentadas em dados concretos.

Diante desse cenário, a identificação precoce de alunos em risco de evasão torna-se uma estratégia importante para a implementação de ações preventivas e personalizadas, visando à permanência e à conclusão bem-sucedida do curso.

Este trabalho tem como objetivo principal aplicar três abordagens de aprendizado de

máquina: a rede neural MLP (*Multilayer Perceptron*), o algoritmo de árvore de decisão *Decision Tree Classifier* e o *Extra Trees Classifier*, para prever, de forma individualizada, a evasão acadêmica nos cursos de TI do IFPE. Utilizando dados acadêmicos e demográficos dos estudantes, busca-se explorar o potencial desses modelos preditivos na tarefa de classificação da evasão, comparando seus desempenhos a fim de identificar quais técnicas oferecem maior precisão e aplicabilidade no apoio à gestão educacional e à tomada de decisões institucionais.

2. Trabalhos Relacionados

Um estudo relacionado foi realizado por (Martinho, Nunes e Minussi, 2013), no qual os autores propuseram um método baseado em redes neurais do tipo Fuzzy ARTMAP para prever o risco de evasão escolar. O trabalho apresentou resultados promissores na identificação de grupos de risco, reforçando a eficácia de abordagens baseadas em aprendizado de máquina para o enfrentamento da evasão no contexto educacional brasileiro.

Outro trabalho relevante foi desenvolvido por (Kawase, 2015), que aplicou modelos de redes neurais do tipo RBF e MLP para analisar a evasão no curso de Sistemas de Informação da UFRRJ. A RBF (*Radial Basis Function*) é uma rede neural que utiliza funções de ativação baseadas na distância entre os dados e pontos centrais, sendo eficaz na identificação de padrões locais. A autora explorou diferentes variáveis acadêmicas e demográficas, demonstrando o potencial dessas técnicas na identificação de perfis com maior probabilidade de evasão.

Um estudo realizado por (Salata, 2019) analisou as razões do abandono escolar entre jovens no Brasil, destacando fatores sociais, econômicos e institucionais que contribuem para a evasão. Embora não tenha utilizado técnicas de aprendizado de máquina, a pesquisa fornece uma base importante para compreender o fenômeno da evasão sob uma perspectiva sociológica, o que complementa abordagens computacionais como as utilizadas neste trabalho.

Este trabalho se diferencia por focar na construção de modelos preditivos voltados à identificação individualizada de estudantes com risco de evasão nos cursos de exatas do IFPE, utilizando dados institucionais e demográficos. Ao contrário de abordagens estatísticas tradicionais, a proposta é oferecer uma aplicação prática que auxilie diretamente a gestão acadêmica, possibilitando intervenções mais rápidas e eficazes. Embora a técnica utilizada se assemelhe aos critérios considerados no cálculo da PNP como a análise de fatores acadêmicos

e demográficos, neste trabalho deu-se ênfase à predição do risco de evasão com base em modelos de aprendizado de máquina, e não ao cálculo direto da evasão em um período específico. Isso reforça o caráter antecipatório da proposta, voltado à tomada de decisão preventiva.

2.1 Aplicação do Classificador *Extra Trees* no Contexto Educacional

O algoritmo Extra Trees Classifier (*Extremely Randomized Trees*) é uma técnica de aprendizado de máquina do tipo ensemble, que constrói múltiplas árvores de decisão a partir de amostras aleatórias dos dados e utiliza critérios aleatórios para as divisões dos nós. Diferentemente de outros métodos baseados em árvores, o Extra Trees seleciona aleatoriamente os pontos de divisão, o que aumenta a diversidade entre as árvores geradas. Essa aleatoriedade contribui para a redução da variância do modelo, tornando-o mais robusto contra o problema de overfitting, que ocorre quando o modelo se ajusta excessivamente aos dados de treinamento, perdendo sua capacidade de generalização. (Geurts et al., 2006).

No contexto educacional, o Extra Trees tem se mostrado uma abordagem promissora em tarefas de predição e apoio à decisão, como a identificação de estudantes em risco de evasão, análise de desempenho acadêmico, classificação de perfis estudantis e elaboração de estratégias institucionais.

Para esta pesquisa, foi utilizada a implementação do algoritmo Extra Trees disponibilizada pela biblioteca *Scikit-learn*, amplamente reconhecida por sua eficiência e facilidade de uso em aplicações de aprendizado supervisionado.

2.2 Aplicação do Classificador *Decision Tree* no Contexto Educacional

O algoritmo *Decision Tree Classifier* (Classificador de Árvore de Decisão) é um modelo de aprendizado supervisionado amplamente utilizado para tarefas de classificação e regressão, devido à sua capacidade interpretativa e à forma intuitiva como estrutura as decisões. Sua lógica consiste em dividir recursivamente o conjunto de dados com base nos atributos mais informativos, formando uma estrutura hierárquica semelhante a uma árvore, com nós de decisão, galhos e folhas representando classes ou valores preditos.

Segundo (Pedregosa et al., 2011), a biblioteca *Scikit-learn* fornece uma implementação eficiente e amplamente utilizada desse algoritmo, adequada para tarefas supervisionadas que exigem interpretabilidade e bom desempenho computacional.

No contexto educacional, sua aplicação é relevante por permitir a visualização clara

das regras de decisão, o que facilita a compreensão dos fatores que mais influenciam o desempenho acadêmico, a evasão escolar e outros fenômenos. Além disso, sua simplicidade computacional e capacidade de lidar com dados heterogêneos a tornam uma alternativa eficiente, especialmente quando a interpretabilidade é uma prioridade.

Neste estudo, foi adotada a versão do *Decision Tree Classifier* disponibilizada pela biblioteca *Scikit-learn*, parametrizada de acordo com os melhores resultados obtidos durante a fase de experimentação.

2.3 Aplicação do Classificador *Multilayer Perceptron* no Contexto Educacional

Optamos por utilizar a rede neural do tipo MLP por sua facilidade de aplicação e sua boa performance com dados não lineares, característica importante ao lidarmos com uma base que contém informações socioeconômicas e demográficas. Para que o modelo fosse capaz de processar corretamente essas variáveis, aplicamos a técnica *One Hot Encoding*, conforme citado anteriormente, convertendo variáveis categóricas em formato numérico.

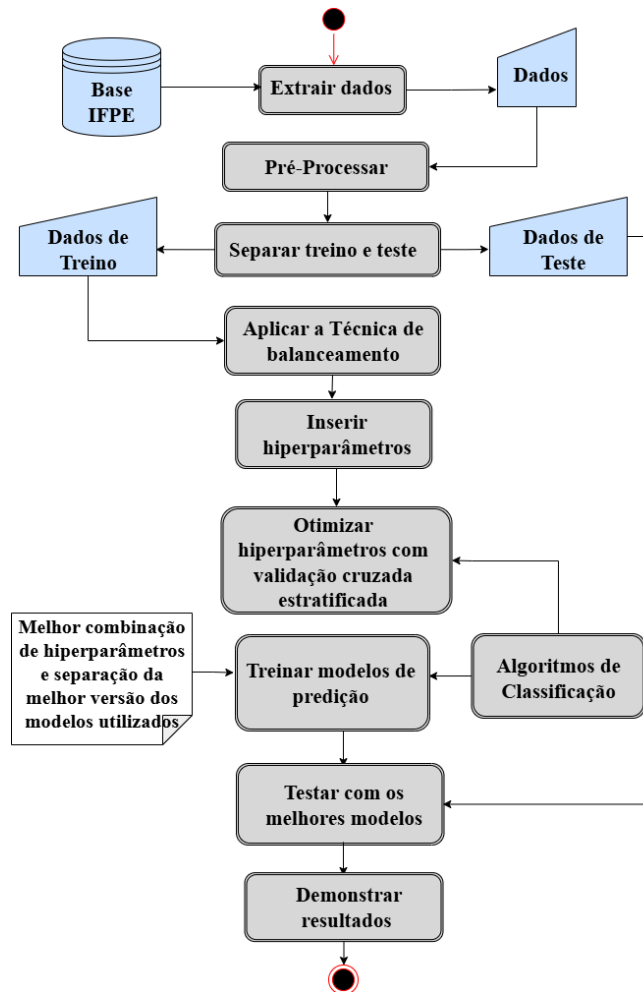
Esse algoritmo é inspirado no funcionamento das redes neurais biológicas, cuja base teórica foi inicialmente proposta por (McCulloch e Pitts, 1943), ao descreverem um modelo lógico para representar a atividade nervosa por meio de neurônios artificiais. O MLP é composto por múltiplas camadas de neurônios interconectados e, durante o treinamento, utiliza a técnica de *backpropagation*, na qual os erros obtidos na saída são retropropagados para ajustar os pesos das conexões sinápticas, promovendo melhorias gradativas no desempenho do modelo.

Dessa forma, a escolha pelo MLP se justifica pela sua versatilidade, pela capacidade de generalização e pelo seu bom desempenho em tarefas supervisionadas com variáveis de alta dimensionalidade e natureza heterogênea, características observadas na base de dados utilizada neste estudo.

3. Metodologia

Para o desenvolvimento deste trabalho, diversos procedimentos foram realizados, abrangendo desde a extração dos dados até o treinamento e a comparação dos modelos preditivos. Na Figura 1, é apresentada uma visão geral da metodologia seguida neste estudo.

Fig. 1. Etapas da metodologia aplicada.



3.1 Conjunto de Dados

Os dados utilizados neste trabalho foram fornecidos pelo Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco (IFPE), abrangendo registros do período entre o semestre 2001.1 e 2025.1, exclusivamente de cursos da área de exatas.

Inicialmente, foi realizada uma análise exploratória da base com o objetivo de compreender e estabelecer relações entre as colunas disponíveis. O conjunto de dados, no formato de CSV, continha 50 colunas e 16.264 linhas.

Para fins de rotulagem, foram considerados alunos evadidos aqueles com situação de matrícula em uma das seguintes categorias: Abandono/Evasão, Cancelamento Voluntário, Jubilado, Cancelamento Compulsório ou Transferido Externo. Já os alunos não evadidos foram identificados pelas categorias: Formado, Concluído e Aguardando Colação de Grau.

Além disso, registros vinculados a situações como Matrícula Vínculo Institucional, Matriculado, Aguardando ENADE, Concludente, Trancado, Estagiário (Concludente),

Transferido Interno, Falecido e Projeto Final (Concludente) foram removidos da base, por representarem alunos com matrícula ainda ativa ou em transição. Como o modelo desenvolvido é supervisionado, é necessário trabalhar apenas com dados rotulados, ou seja, com alunos que já possuem um desfecho conhecido, evasão ou não evasão. No total, a base final contou com 8.727 alunos evadidos e 2.904 alunos não evadidos, totalizando 11.631 registros.

3.2 Técnicas de Pré-processamento e Balanceamento de Dados

Inicialmente, foram removidas colunas que apresentavam identificadores únicos, informações redundantes, alta taxa de valores ausentes ou que poderiam causar vazamento de dados. As colunas Cod_Aluno, cod_matricula e Matrícula foram excluídas por conterem identificadores únicos que não contribuem para a análise preditiva. Colunas como Cod_instituicao, Cod_curso, Turma Atual, Cod_cidade e Nível/Regime de Ensino foram consideradas irrelevantes, pois seus dados já estavam representados em outras colunas com informações equivalentes. A coluna Desc_Matriz_Curricular também foi excluída por sua baixa relevância. Outras colunas foram removidas devido à elevada porcentagem de dados faltantes, como Pólo Municipal (90,76%), necessidades_especiais (98,51%), N_Membros_Familia (88,79%), Dt_Conclusao_Curso (80,10%), Renda Familiar Per Capita PNP (63,42%), Renda Familiar Per Capita SIG (83%), Percentual Frequência (81,77%), Area Procedência Escola Origem (52%) e Grupo Étnico, cuja ausência de dados era praticamente total (99,98%).

A coluna Renda Familiar também foi removida, pois além de apresentar 63,19% de dados ausentes, sua informação já está contemplada pela coluna Renda Familiar Per Capita, que foi mantida no conjunto de dados e fornece uma medida mais normalizada para fins de modelagem. Também foram excluídas colunas que poderiam induzir o modelo a um aprendizado incorreto ou enviesado por vazamento de dados, como Último Evento de Matrícula, Sit. Últ. Per. Letivo, Última presença e Ano_Conclusao_Ensino_Medio, por conterem informações que não estariam disponíveis no momento da previsão.

Além disso, a coluna EIXO_TECNOLÓGICO foi descartada por não oferecer variabilidade útil, visto que todos os cursos pertenciam à área de tecnologia. Por fim, a coluna Situação Matrícula foi retirada do conjunto de entrada por se tratar da variável alvo da análise.

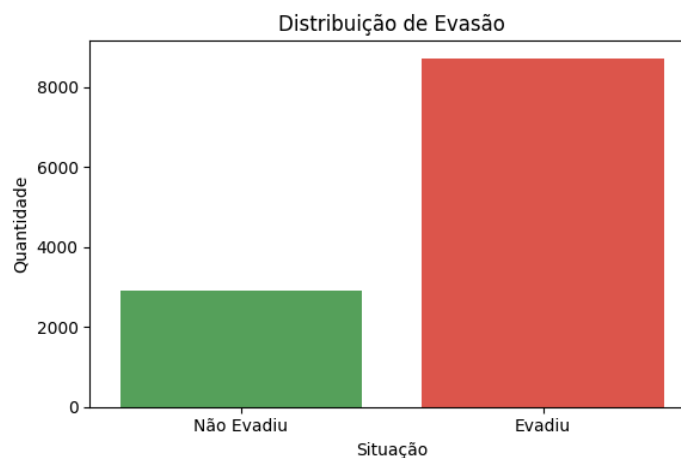
Ao todo, 25 colunas foram excluídas. Após essas remoções e a adição da variável alvo, a base de dados passou a contar com 26 colunas e 11.631 linhas.

Nossa base contempla dados dos seguintes cursos: Análise e Desenvolvimento de

Sistemas, Técnico em Manutenção e Suporte em Informática, Técnico em Informática para Internet, Operador de Computador, Engenharia de Software, Técnico em Informática, Licenciatura em Computação, Técnico em Desenvolvimento de Sistemas, Sistemas para Internet, Técnico em Redes de Computadores e Técnico em Computação Gráfica.

Como mencionado anteriormente, foi realizada a rotulagem dos casos de evasão e não evasão. Após essa etapa e a exclusão das colunas desnecessárias, foi possível visualizar o gráfico de evasão.

Figura 2. Gráfico da distribuição de evasão a partir dos dados extraídos.



Contendo alunos evadidos: 8727 e 2904 não evadidos.

Em seguida, as colunas foram separadas entre categóricas e numéricas, a fim de aplicar técnicas específicas de pré-processamento. Para as colunas numéricas com valores ausentes, utilizou-se a mediana como forma de imputação. Já nas colunas categóricas com dados faltantes, os valores foram substituídos por "Desconhecido", e foi aplicada a técnica de codificação one-hot encoding.

A base de dados foi inicialmente dividida em 80% para treinamento e 20% para teste, respeitando a proporcionalidade das classes por meio da técnica de stratified sampling. Considerando o desbalanceamento presente entre as classes com um número significativamente menor de registros da classe representando a evasão, foi aplicada a técnica *SMOTE* (*Synthetic Minority Over-sampling Technique*) no conjunto de treinamento. O *SMOTE* atua gerando novas amostras sintéticas da classe minoritária com base na interpolação entre exemplos reais vizinhos, evitando a simples duplicação de registros. Essa abordagem busca melhorar a capacidade de generalização dos modelos, prevenindo o viés em favor da classe majoritária e possibilitando que o algoritmo aprenda padrões representativos de ambas as classes de forma equilibrada.

Os hiperparâmetros foram configurados de forma a influenciar diretamente a capacidade de generalização dos modelos. Para realizar esse ajuste de maneira criteriosa, utilizou-se a técnica de validação cruzada estratificada combinada com a busca em grade (*Grid Search*). Nessa abordagem, diversas combinações de configurações dos modelos são testadas e avaliadas por meio de subdivisões do conjunto de dados de treinamento. A validação cruzada estratificada garante que a distribuição das classes seja preservada em cada subdivisão, assegurando uma avaliação mais justa e representativa.

3.3 Métricas de Avaliação

Nesse estudo foram utilizadas métricas de avaliação no desempenho dos modelos preditivos utilizados. Essas métricas possibilitam uma análise detalhada da qualidade das previsões, especialmente em contextos com classes desbalanceadas, como é o caso da evasão acadêmica.

A seguir, os conceitos dos elementos:

- Verdadeiro Positivo (TP): quando o modelo identifica corretamente um estudante que evadiu.
- Verdadeiro Negativo (TN): quando o modelo acertou ao prever que o estudante permaneceu.
- Falso Positivo (FP): quando o modelo prevê evasão, mas o estudante permanece.
- Falso Negativo (FN): quando o modelo não detecta a evasão, mas o estudante efetivamente desiste.

A partir dessas definições, é possível derivar indicadores estatísticos fundamentais para a avaliação do desempenho dos modelos preditivos. As métricas foram:

- Acurácia: expressa a proporção total de acertos, considerando tanto as previsões corretas de evasão quanto de permanência, em relação ao total de casos avaliados.
- Precisão: quantifica a proporção de acertos entre todas as previsões positivas realizadas, indicando o quanto as previsões de evasão são confiáveis.
- *Recall*: mensura a capacidade do modelo em identificar corretamente os estudantes que realmente evadiram, sendo crucial em contextos onde se busca minimizar falsos negativos.
- F1-Score: Corresponde à média harmônica entre precisão e recall, funcionando como uma métrica balanceada para cenários em que há desequilíbrio entre as classes, como é comum em bases com baixa taxa de evasão.

Essas métricas oferecem uma visão abrangente do desempenho dos modelos, equilibrando a taxa de acertos, a confiabilidade das predições e a detecção efetiva dos estudantes em risco (SWAMINATHAN; TANTRI, 2024).

3.4 Comparativo dos Modelos

Conforme citado nas seções 2.1, 2.1 e 2.3, os classificadores *Extra Trees*, *Decision Tree* e MLP possuem características distintas que os tornam adequados para a tarefa de predição da evasão acadêmica, a depender do contexto e dos objetivos da aplicação.

O *Extra Trees Classifier*, abordado na seção 2.1, mantém a base interpretável dos modelos de árvore, mas se diferencia por incorporar aleatoriedade tanto na seleção das amostras quanto nas divisões dos nós. Essa característica torna o modelo mais robusto, favorecendo uma melhor capacidade de generalização, o que o torna uma opção interessante quando se busca equilíbrio entre interpretabilidade e desempenho.

O *Decision Tree Classifier*, descrito na seção 2.2, é reconhecido por sua simplicidade e interpretabilidade. Sua estrutura em forma de árvore facilita a visualização das regras de decisão, o que permite compreender com clareza os fatores que influenciam a evasão. Por essa razão, é especialmente útil em cenários onde a transparência do modelo é fundamental. No entanto, sua estrutura única o torna mais suscetível ao superajuste, especialmente em bases complexas.

O MLP, descrito na seção 2.3, é uma rede neural de múltiplas camadas capaz de modelar relações não lineares entre variáveis. Sua aplicação é especialmente recomendada em conjuntos de dados com alta dimensionalidade e natureza heterogênea, como ocorre em bases educacionais que envolvem informações acadêmicas, demográficas e socioeconômicas. Apesar da menor interpretabilidade em relação aos modelos baseados em árvores, o MLP oferece maior flexibilidade na modelagem de padrões complexos.

Cada técnica apresenta vantagens específicas, por isso, a escolha entre elas é levada em consideração o equilíbrio desejado entre transparência, capacidade de generalização e complexidade dos dados.

4. Resultados

Após a realização do treinamento e ajuste dos modelos por meio da técnica de validação cruzada e a utilização do *Grid Search*, foram obtidos os valores das principais métricas de avaliação: acurácia, precisão, *recall* e F1-score. Essas métricas foram calculadas

com base nos dados de teste, previamente separados, e têm como objetivo mensurar o desempenho dos algoritmos na tarefa de prever a evasão acadêmica.

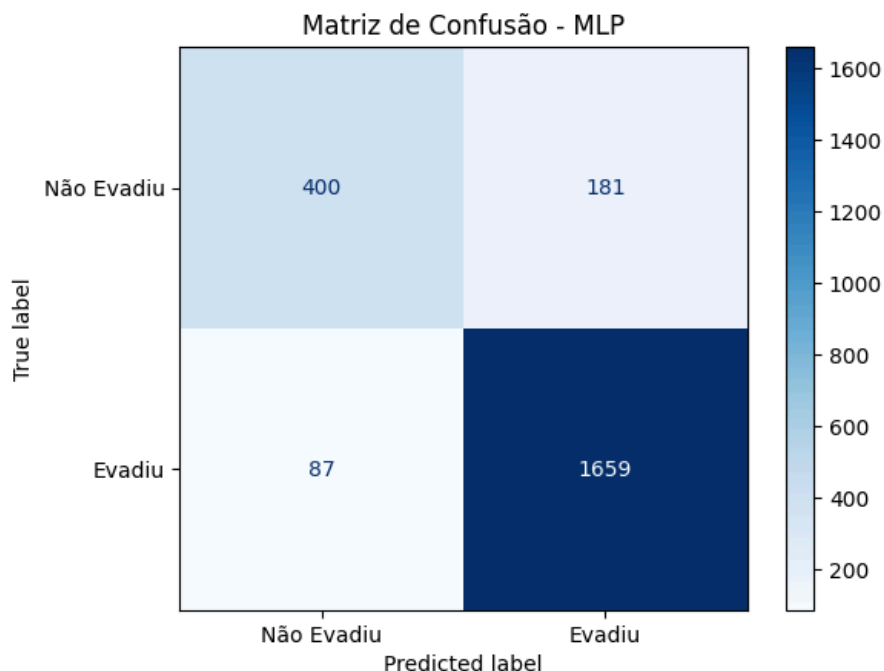
Esses resultados reforçam a importância de alinhar a escolha dos modelos às características da base de dados e aos objetivos da tarefa. Em cenários como o da evasão acadêmica, com dados heterogêneos e complexos, modelos capazes de capturar padrões não lineares e sutis tendem a apresentar melhor desempenho preditivo.

Tabela 1. Resultados a partir das métricas de avaliação utilizadas.

Algoritmos	<i>FI_Score</i>	Precisão	Revocação (Recall)	Acurácia
MLP	92%	90%	95%	88,5%
<i>Decision Tree Classifier</i>	90%	90%	90%	85,5%
<i>Extra Trees Classifier</i>	83%	87%	80%	76,9%

A seguir, são apresentadas as matrizes de confusão de cada modelo, acompanhadas da análise de seu desempenho.

Figura 3. Matriz de confusão do modelo M

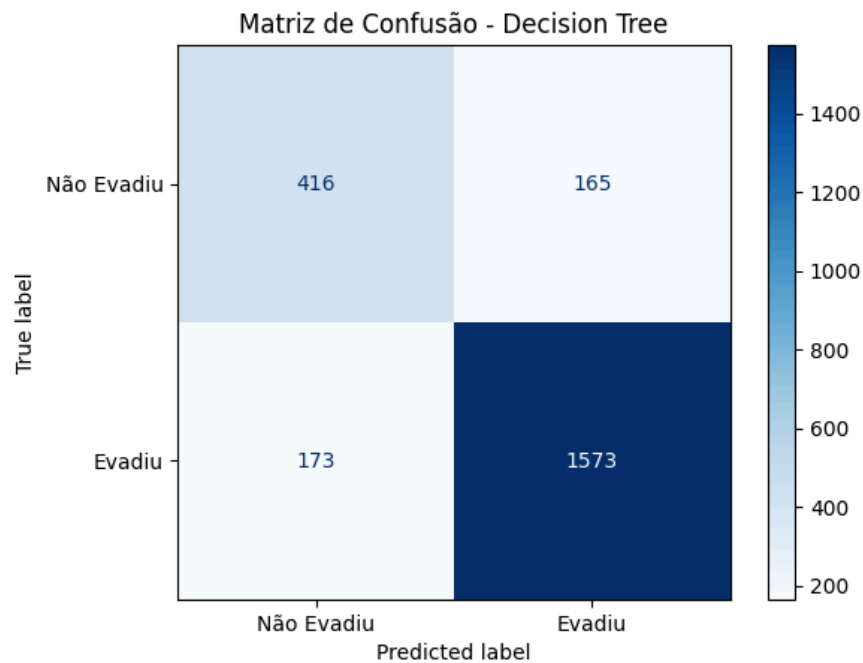


- 1659 acertos - ao classificar que os estudantes evadiram (Verdadeiros Positivos)
- 181 erros - na previsão de evasão para estudantes que permaneceram no cursos (Falsos Positivos)

- 400 acertos - na previsão de alunos que permaneceram (Verdadeiros Negativos)
- 87 erros - ao classificar que permaneceram estudantes que na verdade evadiram (Falsos Negativos)

A taxa baixa de falsos negativos é particularmente relevante do ponto educacional, pois demonstra que o modelo é capaz de identificar a maioria dos alunos em risco de evasão.

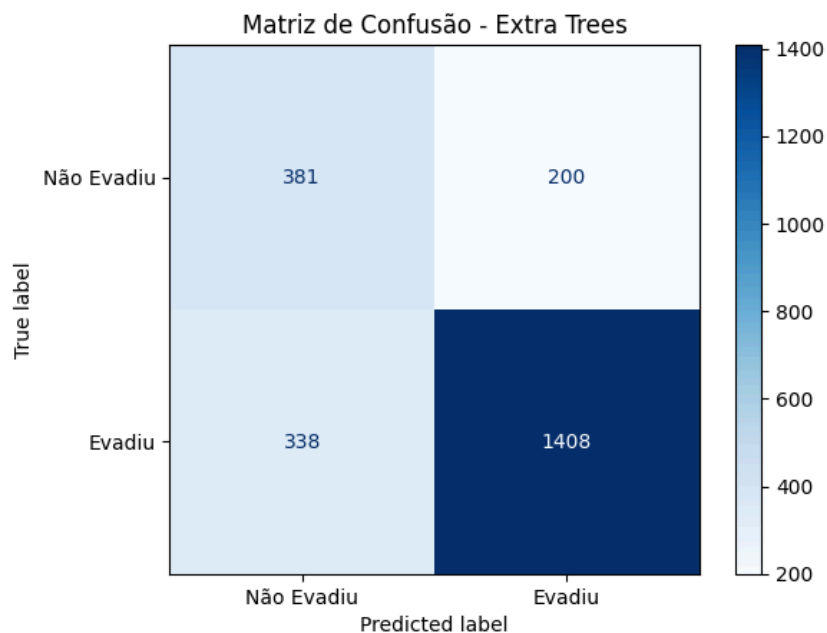
Figura 4. A matriz de confusão referente ao modelo *Decision Tree*



- 1573 acertos - ao classificar que os estudantes evadiram (Verdadeiros positivos)
- 165 erros - na previsão de evasão para estudantes que permaneceram no curso (Falso positivos)
- 416 acertos - na previsão de alunos que permaneceram (Verdadeiros Negativos)
- 173 erros - ao classificar que permaneceram estudantes que na verdade evadiram (Falsos Negativos)

O resultado da matriz de confusão tem uma boa taxa de acerto, mas ainda há espaço para melhorias na redução de falsos negativos, pois esses representam os estudantes em risco de evasão não identificados corretamente.

Figura 5. A matriz de confusão referente ao *Extra Trees*



- 1408 acertos - ao classificar que os estudantes evadiram (Verdadeiro positivo)
- 200 erros - na previsão de evasão para estudantes que permaneceram no curso (Falso positivo)
- 381 acertos - ao classificar que os estudantes permaneceram (Verdadeiros negativos)
- 338 erros - ao classificar que permaneceram estudantes que na verdade evadiram (Falsos negativos)

O número elevado de falsos negativos indica uma limitação importante do modelo em identificar estudantes em risco real de evasão. A partir da maior taxa de erro nessa classe, percebemos que o modelo tem menos sensibilidade (*Recall*) em comparação aos demais modelos.

5. Conclusão

Este estudo utilizou técnicas de aprendizado de máquina para a predição da evasão acadêmica em cursos da área de Exatas do Instituto Federal de Pernambuco (IFPE). A análise foi realizada a partir de dados acadêmicos, demográficos e socioeconômicos, com o objetivo de identificar padrões e fatores que influenciam o risco de abandono escolar.

Os resultados indicaram que o modelo de rede neural MLP apresentou maior capacidade preditiva, com taxa de acerto (acurácia) de 88,5%, precisão de 90%, *recall* de 95% e F1-Score de 92%. O modelo se destacou especialmente na identificação dos estudantes com

maior probabilidade de evasão, devido à sua habilidade de lidar com relações complexas e não lineares entre as variáveis. O modelo *Decision Tree Classifier*, apesar de apresentar desempenho inferior, possibilitou uma melhor compreensão dos principais fatores associados à evasão, como o desempenho acadêmico e condições socioeconômicas.

Embora o estudo não tenha explorado diretamente desigualdades sociais mais amplas, a aplicação de modelos preditivos neste contexto pode contribuir significativamente para o desenvolvimento de políticas educacionais mais eficazes e para a redução dos índices de evasão no IFPE.

6. Referências

GONZALEZ-NUCAMENDI, Andres; NOGUEZ, Julieta; NERI, Luis; ROBLEDO-RELLA, Víctor; GARCÍA-CASTELÁN, Rosa María Guadalupe. Predictive analytics study to determine undergraduate students at risk of dropout. *Frontiers in Education*, v. 8, 2023.

KAWASE, Kelly Harumi Fausta. Aplicação de Redes Neurais RBF e MLP na análise de evasão discente do curso de Sistemas de Informação da UFRRJ. 2015. 125 f. Dissertação (Mestrado em Ciências) – Universidade Federal Rural do Rio de Janeiro, Instituto de Ciências Exatas, Programa de Pós-Graduação em Modelagem Matemática e Computacional, Seropédica, RJ, 2015.

LOPES DA SILVA GARCIA, L. M.; GOMES, Raquel Salcedo. Causas da evasão em cursos de ciências exatas: uma revisão da produção acadêmica. *Revista Educar Mais*, [S. l.], v. 6, p. 937–957, 2022. DOI: 10.15536/reducarmais.6.2022.2970. Disponível em: <https://periodicos.ifsul.edu.br/index.php/educarmais/article/view/2970>. Acesso em: 21 fev. 2025.

MARTINHO, Valquíria R. C.; NUNES, Clodoaldo; MINUSSI, Carlos R. A New Method for Prediction of School Dropout Risk Group Using Neural Network Fuzzy ARTMAP. In: *INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE (ICAI)*. Proceedings [...]. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2013.

MCCULLOCH, Warren S.; PITTS, Walter. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, v. 5, p. 115–133, 1943.

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

SALATA, André. Razões da evasão: abandono escolar entre jovens no Brasil. *Interseções: Revista de Estudos Interdisciplinares*, Rio de Janeiro, v. 21, n. 1, 2019. Disponível em: <https://journals.openedition.org/intersecoes/5569>. Acesso em: 11 jun. 2025.

SENA, G. et al. SIOG2024-1-OA-004 Predicting early mortality in older adults with stomach cancer: a machine learning (ML) Approach incorporating comprehensive geriatric assessment (CGA) and clinical data (CD). *Journal of Geriatric Oncology*, v. 15, n. 7, p. 101848, 2024.

SOUZA, Elisângela de; FREITAS, Lourdes Francisca. Um estudo sobre a evasão nos cursos de graduação dos Institutos Federais. *Revista Brasileira da Educação Profissional e Tecnológica*, [S. l.], v. 1, n. 20, p. e10757, 2021. DOI: 10.15628/rbept.2021.10757.

SWAMINATHAN, S.; TANTRI, B. R. Confusion matrix–based performance evaluation metrics. *African Journal of Biomedical Research*, v. 27, n. 4S, p. 4023–4031, 2024.

SERVIÇO PÚBLICO FEDERAL
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE PERNAMBUCO
BIBLIOTECA DEPOSITÁRIA – CAMPUS PALMARES

TERMO DE AUTORIZAÇÃO PARA DISPONIBILIDADE DE TCC NO REPOSITÓRIO INSTITUCIONAL IFPE

Na qualidade de titular dos direitos de autor da publicação, autorizo ao Repositório Institucional do Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco (RIIFPE) a disponibilizar, através do site <https://repositorio.ifpe.edu.br/xmlui/>, sem ressarcimento dos direitos autorais, de acordo com a Lei n. 9.610/98, o texto integral da obra abaixo citada, a título de divulgação e de preservação digital da produção científica brasileira, a partir desta data.

Identificação:

Autor	Yanka Mirelly Ayala de Sales Silva
E-mail	ymass@discente.ifpe.edu.br
Link Lattes	https://lattes.cnpq.br/1256933905965539
Orcid	https://orcid.org/0009-0008-7543-5072
Título	Aplicação de Modelos de Aprendizado de Máquina na Predição da Desistência Acadêmica em Cursos de TI do IFPE
Orientador(a)	Verlayne Kelley da Hora Rocha Araujo
Data de defesa	25/07/2025
ODS Agenda 2030 (quando aplicável)	ODS 4 – Educação de Qualidade

LICENÇA DE DIREITO AUTORAL

Na qualidade de titular dos direitos de autor do conteúdo supracitado, autorizo o Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco a disponibilizar a obra, gratuitamente, de acordo com a licença pública Creative Commons, Licença 4.0 Unported por mim declarada sob as seguintes condições:

Permitir uso comercial da obra? Sim Não

Permitir modificações em sua obra?


Sim

Sim, contanto que outros compartilhem pela mesma licença

Não

A obra continua protegida por direito autoral e/ou por outras leis aplicáveis, respeitando inclusive o contrato celebrado entre a editora ou periódico que veicula a mesma. Qualquer uso da obra que não o autorizado sob esta licença ou pela legislação autoral é proibido.

Documento assinado digitalmente

 **YANKA MIRELLY AYALA DE SALES SILVA**
Data: 12/01/2026 18:21:37-0300
Verifique em <https://validar.iti.gov.br>

12/01/2026


Data de autorização

SERVIÇO PÚBLICO FEDERAL
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE PERNAMBUCO
BIBLIOTECA DEPOSITÁRIA – CAMPUS PALMARES

TERMO DE RESPONSABILIDADE DO ORIENTADOR E AUTORIZAÇÃO PARA PUBLICAÇÃO DA PRODUÇÃO INTELECTUAL NO REPOSITÓRIO INSTITUCIONAL RI-IFPE

Eu, Verlaynne Kelley da Hora Rocha Araujo, CPF:037695744-19 , ORCID:0009-0008-0770-9406 , link lattes:<http://lattes.cnpq.br/2991700693568241>, orientadora do Trabalho de Conclusão de Curso, intitulado: Aplicação de Modelos de Aprendizado de Máquina na Predição da Desistência Acadêmica em Cursos de TI do IFPE, realizado pelo aluno(a): Yanka Mirelly Ayala de Sales Silva, Matrícula nº 20222ADS-PM0369, Curso: Análise e Desenvolvimento de Sistemas, autorizo o depósito obrigatório, na Biblioteca do Campus Palmares, da versão final, no modelo de: (x) artigo () monografia () dissertação ().

INFORMAÇÕES DA BANCA			
NOME	CPF	ORCID	LINK LATTES
Verlaynne Kelley da Hora Rocha Araujo	037695744-19	0009-0008-0770-9406	http://lattes.cnpq.br/2991700693568241
Maria Carolina Torres da Silva	062.858.264-12	https://orcid.org/0000-0002-5968-6495	http://lattes.cnpq.br/6577076443532261
Thaise Cristina da Silva Neri	089439854-79		http://lattes.cnpq.br/4537732201212377

Documento assinado digitalmente
 **VERLAYNNE KELLEY DA HORA ROCHA ARAUJO**
Data: 12/01/2026 19:01:24-0300
Verifique em <https://validar.iti.gov.br>

Professora orientadora