

Identificação, Análise e Apoio à Tomada de Decisão sobre Evasão Escolar no Contexto de um Curso de Engenharia de Computação

Rafaela Oliveira Lorenzeto
Braga
Centro Federal de Educação
Tecnológica de Minas Gerais
Leopoldina, MG, Brasil
rafaelaobraga@hotmail.com

Humberto Motta da Cunha
Centro Federal de Educação
Tecnológica de Minas Gerais
Leopoldina, MG, Brasil
humbertomottadacunha@gmail.com

Gabriella Castro Barbosa Costa
Dalpra
Centro Federal de Educação
Tecnológica de Minas Gerais
Leopoldina, MG, Brasil
gabriella@cefetmg.br

Luís Augusto Mattos Mendes
Centro Federal de Educação
Tecnológica de Minas Gerais
Leopoldina, MG, Brasil
luisaugusto@cefetmg.br

Luan Soares Oliveira
Centro Federal de Educação
Tecnológica de Minas Gerais
Leopoldina, MG, Brasil
luan@cefetmg.br

ABSTRACT

The school dropout rate in the IT field in Brazil was around 60.2% in 2017, according to data released by the Ministry of Education's Higher Education Census. Considering this fact, this paper aims to establish patterns regarding students who have abandoned the Computer Engineering course at a federal institution in Brazil, using a predictive model. An intelligent system capable of predicting if a student will drop out or not based on his academic history was proposed and implemented. The methodology for the system development included four distinct phases. In the first phase, a data extraction algorithm from academic histories was developed. The second phase encompassed correlation analyses and the generation of a decision tree to identify dropout indicators, using data from 180 students. Despite data limitations, the model achieved 78% accuracy. In the third and fourth phase an Application Programming Interface (API) and a system dashboard were implemented to achieve the proposed goal.

KEYWORDS

University Dropout, Computer Engineering, Decision Tree

1 INTRODUÇÃO

O Art. 43 da LDB (Lei de Diretrizes e Bases da Educação) de 1996 [1] evidencia que a educação superior tem por finalidade estimular a criação cultural, o desenvolvimento do espírito científico e do pensamento reflexivo, formando diplomados em diversas áreas de conhecimento. Esses profissionais são preparados para inserção em setores profissionais e para contribuir no desenvolvimento da sociedade brasileira, colaborando também em sua formação contínua. Além disso, a educação superior busca incentivar o trabalho de pesquisa e investigação científica, visando o avanço da ciência, tecnologia e a difusão da cultura, contribuindo para o entendimento do ser humano e do meio em que vive. Dessa forma, temos que as instituições de ensino superior possuem um papel fundamental na formação profissional e pessoal de um indivíduo, desempenhando um papel crucial na edificação de uma sociedade mais equitativa. Elas exercem um papel ativo na disseminação e produção de conhecimento, além de impulsionar o desenvolvimento tecnológico,

econômico e social. À medida que mais estudantes concluem seus cursos e ingressam no mercado de trabalho como profissionais capacitados, a contribuição social das universidades se amplia.

Apesar do Art. 205 da Constituição Brasileira [2] garantir que “a educação, direito de todos e dever do Estado e da família, será promovida e incentivada com a colaboração da sociedade, visando ao pleno desenvolvimento da pessoa, seu preparo para o exercício da cidadania e sua qualificação para o trabalho”, constata-se que a evasão escolar vem aumentando ao decorrer de cada dia. Conforme levantamento realizado pelo Datafolha a pedido do C6 Bank, 8,4% dos estudantes (cerca de 4 milhões) com idade entre 6 e 34 anos abandonaram a escola durante a pandemia de COVID-19. Já a taxa de abandono no âmbito do ensino superior é de cerca de 16,3% dos estudantes [3]. Ademais, a divulgação do Censo da Educação Superior do Ministério da Educação, feita pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, relata que a evasão do curso nas áreas de computação atingiu o índice de 60,2% em 2017 [4].

É caracterizado como evasão escolar quando falamos de um aluno que não retornou à instituição de ensino para dar continuidade em seu aprendizado ou não fez sua matrícula. Pode-se distinguir a evasão escolar no ensino superior em três tipos: (i) a **evasão de curso**, que é quando o estudante deixa de matricular-se, desiste ou há transferência de curso; (ii) a **evasão da instituição**, que ocorre quando o estudante desliga-se da instituição na qual está matriculado e, por último, (iii) a **evasão do sistema**, que caracteriza o abandono de forma definitiva ou temporária ao ensino superior [5].

Tem-se como contexto para a realização deste trabalho a **evasão de curso** do curso de Engenharia de Computação, ofertado pelo Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG), campus Leopoldina, desde o segundo semestre letivo do ano de 2018. O presente trabalho tem como objetivo traçar relações entre dados de alunos que evadiram do curso, ou seja, que realizaram o cancelamento de suas matrículas ou tiveram suas matrículas alteradas para o status de cancelada e os alunos que ainda estão ativos no curso de Engenharia de Computação. Por meio da extração automatizada dos dados dos históricos de 180 alunos e da

utilização de um modelo preditivo baseado em árvore de decisão, foi possível traçar um perfil / alcançar informações baseadas nos dados acadêmicos destes alunos, e, com isso, prever futuros estudantes propensos à evasão. Com a obtenção desse perfil e através da utilização do sistema inteligente desenvolvido, torna-se possível a realização de ações preventivas visando evitar ou minimizar a evasão escolar no contexto apresentado.

O restante deste trabalho está organizado da seguinte forma: na Seção 2 são apresentados trabalhos relacionados a esta proposta; na Seção 3 tem-se o detalhamento do trabalho desenvolvido, seguida da Seção de Resultados. Por fim tem-se, na Seção 5, as Considerações Finais, seguida dos Agradecimentos e Referências utilizadas.

2 TRABALHOS RELACIONADOS

Uma diversidade de pesquisas tem se voltado para estratégias visando diminuir a evasão escolar nas instituições, explorando uma gama de áreas acadêmicas e sugerindo medidas para atenuar esse fenômeno, além de explorar os motivos que levam um estudante a evadir de sua instituição de ensino. Os estudos citados nesta seção apresentam similaridades nesse contexto, pois todos se concentram no ensino superior na área da Computação; contudo, cada um se distingue ao analisar diferentes causas de evasão, adotar abordagens de avaliação específicas e situar-se em contextos particulares.

Uma pesquisa conduzida no curso de Bacharelado em Ciência da Computação da Universidade de Brasília [6] se propôs a investigar os fatores que contribuem para a alta taxa de evasão no curso, que ultrapassa os 50%. Utilizando dados pessoais e acadêmicos dos alunos fornecidos pelo sistema acadêmico da instituição, o objetivo foi identificar os principais elementos ligados ao perfil dos estudantes em risco e, assim, encontrar estratégias para mitigar esse problema. A análise estatística revelou que a reprovação em disciplinas obrigatórias, especialmente devido a trancamentos ou baixo rendimento, está correlacionada com a evasão. Diversos classificadores foram desenvolvidos, destacando-se o *Naive Bayes* como o mais eficaz. Segundo esse classificador, cerca de 58.56% dos alunos atualmente matriculados não concluirão o curso. Com a taxa de evasão atual em 55.76%, evidencia-se uma preocupante tendência de agravamento do problema.

Por sua vez, em um estudo sobre evasão no curso de Engenharia de Computação na UEFS [7], a metodologia utilizada foi baseada em abordagens qualitativas e quantitativas, onde foram aplicadas técnicas mistas, incluindo questionários, entrevistas e pesquisa documental, a fim de investigar as taxas de evasão, causas e motivos da desistência dos alunos. A pesquisa contou com a participação de 53 alunos evadidos, sendo 86,79% do sexo masculino e 13,20% do sexo feminino, com uma média de idade de 21,6 anos e tempo médio de permanência no curso de 4,3 semestres. A partir da aplicação das técnicas citadas anteriormente (questionários, entrevistas semi-estruturadas e pesquisa documental), foram obtidos os seguintes resultados: o questionário identificou que as principais razões da evasão foram “falta de motivação”, “dificuldade de adaptação ao método PBL (*Problem-Based Learning*)” e “insatisfação com o curso” como motivos mais citados pelos alunos evadidos. Já as entrevistas aprofundaram as motivações da evasão, buscando captar significados que não seriam alcançados apenas com o questionário. A pesquisa documental teve o objetivo de levantar dados concretos

para calcular a evasão do curso. Os resultados do estudo revelaram que os alunos que evadiram com mais de quatro semestres apresentaram maior dificuldade em conciliar estudo e trabalho, enquanto os que evadiram nos primeiros semestres destacaram como motivos principais não terem escolhido o curso como primeira opção e dificuldades de adaptação ao método de ensino. Essas diferenças apontaram para a importância de medidas de prevenção da evasão e para a necessidade de acompanhamento mais próximo dos alunos em situação de risco. Portanto, o estudo forneceu informações relevantes para a compreensão da evasão no curso de Engenharia de Computação na UEFS e sugestões para a promoção da permanência dos estudantes.

Já no curso de bacharelado em Ciência da Computação da UFERSA, campus Mossoró/RN [8], o estudo teve como objetivo identificar os motivos por trás das altas taxas de evasão acadêmica. A base de dados foi extraída do Sistema Integrado de Gestão de Atividades Acadêmicas (SIGAA) da instituição, fornecendo informações cruciais sobre o percurso dos alunos antes da evasão. Foram analisados dados de 450 alunos, dos quais 60% evadiram e 40% concluíram o curso. O estudo foi caracterizado como descritivo e aplicado, validado por meio de estudo de caso. A abordagem da pesquisa foi predominantemente quantitativa, com a coleta de dados primários relacionados aos fatores e índices de evasão acadêmica do curso. As variáveis exploradas abrangeram a situação do aluno, residência na vila acadêmica, experiência prévia na área, cor da pele, estado de origem, tipo de escola de ensino médio, forma de ingresso, índice de rendimento acadêmico, total de reprovações e renda familiar. A análise resultou na identificação de dois grupos distintos: alunos que concluíram o curso e aqueles que evadiram. Fatores como cor da pele, origem socioeconômica e relações familiares foram destacados como influenciadores na evasão acadêmica.

3 SOLUÇÃO PROPOSTA

Como primeiro passo para o desenvolvimento da solução acerca da problemática da evasão escolar no contexto do curso de Engenharia de Computação do CEFET-MG, campus Leopoldina, foi desenvolvido um algoritmo para ler e extrair dados dos históricos dos alunos. Em seguida, um segundo algoritmo capaz de classificar os alunos propensos à evasão foi desenvolvido. Como terceiro passo implementou-se uma API (*Application Programming Interface* - Interface de Programação de Aplicações) e, no quarto e último passo, um *dashboard*. Cada um destes passos é apresentado em detalhes a seguir.

Com relação ao algoritmo criado¹, este foi desenvolvido utilizando a linguagem Python² e é capaz de ler os históricos escolares fornecidos pelo SIGAA³, que é Sistema Integrado de Gestão de Atividades Acadêmicas utilizado pela instituição. A partir do cabeçalho do histórico foram extraídos os seguintes dados: valor total de horas cursadas, ano e período que o aluno entrou e/ou saiu da instituição, tipo de saída (se ocorreu), se houve trancamento de matrícula, quantidade de semestres de trancamento, gênero do aluno e o tipo de escola em que foi concluído o ensino médio. Do restante do histórico foram extraídos dados referentes à quantidade

¹https://github.com/florenzeto/iniciacao_cientifica_leitor_de_historico_2022.

²<https://www.python.org/>

³<https://sig.cefetmg.br/sigaa>

de aprovações e reprovações, média de todas as notas e frequência. A leitura dos dados contidos no histórico foi realizada utilizando a biblioteca PyPDF2⁴, enquanto a leitura e extração das tabelas contidas no histórico foram feitas por meio da biblioteca Tabula⁵. Posteriormente, todos os dados foram organizados em um *dataframe* utilizando a biblioteca pandas⁶. Para garantir a limpeza e padronização dos dados, foram aplicados recursos desta mesma biblioteca. Como resultado, os históricos processados pelo algoritmo geraram um *dataframe* no formato de uma tabela do Excel, a ser utilizado na etapa de análise dos dados. Para validar o algoritmo criado, o estudante responsável por sua criação utilizou seu próprio histórico acadêmico.

Após implementar o algoritmo mencionado anteriormente, iniciou-se a manipulação dos dados extraídos por ele, realizando o tratamento desses dados, conduzindo análises sobre as informações processadas e aplicando técnicas estatísticas e preditivas. Após identificar as correlações entre características específicas e o cancelamento de matrículas de alunos, avançou-se para a criação de um algoritmo preditivo capaz de classificar os alunos propensos à evasão. Esse algoritmo também foi desenvolvido utilizando a linguagem de programação Python, com o auxílio das bibliotecas scikit-learn⁷, pandas e numpy⁸.

Como terceiro passo, realizou-se o desenvolvimento de uma API em Python. Essa API foi projetada para receber um modelo preditivo juntamente com os dados de um aluno específico para o qual deseja-se prever sua possibilidade de evasão. A API produz uma classificação para o aluno, indicando se ele se encontra na categoria daqueles cuja matrícula será cancelada ou permanecerá ativa.

Com base nas funcionalidades previamente implementadas e como último passo da solução proposta, avançou-se para a etapa de desenvolvimento de um *dashboard*⁹, com o objetivo de aprimorar a interação do usuário final com os algoritmos desenvolvidos. A construção do *frontend* deste *dashboard* foi conduzida utilizando a linguagem de programação JavaScript com a biblioteca React¹⁰ e folhas de estilo em CSS¹¹. A interface foi estruturada em três partes distintas: a página inicial (Home), que expõe o propósito da interface e sua finalidade; página com informações sobre a equipe e os contatos dos participantes e, por fim, a página de login, onde a API do modelo preditivo está integrada como uma interface de processamento de aplicações entre um servidor da Web e um navegador da Web. A autenticação por meio de login possibilita o acesso ao modelo preditivo, sendo este restrito aos usuários previamente autorizados pelos autores do trabalho. Por fim, a predição é realizada ao preencher os campos requisitados relacionados ao aluno em questão. Após o preenchimento desses dados, a interface se comunica com a API implementada em Python, com suporte das bibliotecas Flask¹², Joblib¹³ e Scikit-Learn. Essa API retorna ao

dashboard a classe à qual o aluno pertence, podendo variar entre "CANCELADO" e "ATIVO".

4 RESULTADOS

Como principal resultado do trabalho proposto tem-se o modelo preditivo baseado em árvore de decisão criado. Para a sua geração, foram realizadas uma série de análises, anteriormente à criação da árvore de decisão e do *dashboard* que permite a sua utilização pelo usuário final. A subseção a seguir descreve cada uma das análises realizadas, relatando dados e métodos utilizados e resultados obtidos. Na Subseção 4.2 é apresentada a árvore de decisão e, por fim, na Subseção 4.3 tem-se o detalhamento do *dashboard* implementado

4.1 Análise

Com intuito de encontrar relações entre o perfil dos próprios alunos cancelados, a primeira análise realizada foi utilizado um total de 29 alunos cujas matrículas foram canceladas, se baseando em relações do tipo de ingresso e tipo de saída (motivo do cancelamento), com o tipo de ingresso tendo sua variação entre pública e particular, e o tipo de saída variando em: solicitação espontânea que é quando o próprio aluno solicita o cancelamento, abandono / desistência que é quando o aluno não realiza sua matrícula por dois semestres consecutivos, e efetivação de novo cadastro que é quando o aluno muda de curso na mesma instituição. Nos dados apresentados foi aplicado a distribuição de qui-quadrado e adotou-se que para valores de probabilidade maiores de 0.05, não haveria correlação. Realizando o teste de independência das variáveis, obtivemos um valor de 0.007441, portanto não encontramos correlação entre as variáveis apresentadas. Para realizar o teste foi utilizado um algoritmo em Python com o auxílio das bibliotecas pandas, numpy e stats models¹⁴. Ademais, vale ressaltar que todos os testes realizados envolvendo a distribuição de qui-quadrado utilizaram o mesmo algoritmo.

Na segunda análise realizada, houve apenas um aumento de dados, subindo de 29 alunos cancelados para 36 alunos. Aplicando o mesmo teste da análise inicial e se baseando nas mesmas relações, obteve-se um valor de probabilidade de 0.007367, contudo ainda não houve correlação, porém houve redução.

Partindo para terceira análise, utilizou-se um total de 36 alunos cuja matrícula fora cancelada e permaneceu-se com o método aplicado anteriormente (distribuição de qui-quadrado), havendo apenas a mudança dos dados de entrada. Os dados utilizados para a terceira análise foram: o gênero, classificado em masculino e feminino e tipo de entrada, pública ou particular. Partindo dos dados mencionados foi aplicado a distribuição de qui-quadrado e obteve-se um valor de probabilidade de 0.1772. Apesar da redução do índice de probabilidade, não foi possível afirmar uma correlação.

Já na quarta análise, adotou-se o mesmo método das análises anteriores e um total de 36 alunos com a matrícula cancelada. Desta vez, aferiu-se a relação entre os dados tipo de saída (motivo de cancelamento) e gênero. Obteve-se um valor de probabilidade de 0.3007, portanto, não foi encontrado relação entre as variáveis apresentadas até então.

Na quinta análise foi utilizado um total de 36 alunos cujas matrículas foram canceladas e os seguintes dados: Total de Semestre,

⁴<https://pypi.org/project/PyPDF2/>

⁵<https://pypi.org/project/tabula-py/>

⁶<https://pandas.pydata.org/>

⁷<https://scikit-learn.org/stable/>

⁸<https://numpy.org/>

⁹<https://main-evasaoescolar.netlify.app/>

¹⁰<https://react.dev/>

¹¹<https://www.w3.org/Style/CSS/>

¹²<https://flask.palletsprojects.com/en/3.0.x/>

¹³<https://pypi.org/project/joblib/>

¹⁴<https://www.statsmodels.org/stable/index.html>

Número de reprovação por nota, Número de aprovação, Semestre sem disciplinas, Número de reprovação por falta, Total de horas realizadas, Total de disciplinas com dispensa, Semestre de entrada e Semestre de saída. Para esta análise foi desenvolvido um novo algoritmo em Python baseado no método de correlação de Pearson. Para a criação deste algoritmo foram utilizadas as bibliotecas pandas, mumpy e scikit-learn. No método de correlação de Pearson verifica-se a correlação de duas variáveis e, para isso, adotou-se que o módulo dos valores denominados Coeficiente de Pearson mais próximos de 1.00 teriam maior correlação. Após a execução do algoritmo foram obtidas as correlações listadas na Tabela 1.

Variável 1	Variável 2	Coef. Pearson
Total de semestre	Semestre sem disciplina	0.61
Número de reprovação por nota	Semestre sem disciplina	0.77
Número de reprovação por nota	Número de reprovação por falta	0.78
Número de aprovação	Total de horas realizadas	0.90
Total de disciplinas com dispensa	Número de aprovação	0.66
Total de disciplinas com dispensa	Total de horas realizadas	0.91
Total de semestre	Número de aprovação	0.71
Semestre de entrada	Semestre de saída	0.81

Tabela 1: Correlação de Person

Considerando os resultados apresentados na Tabela 1, foi realizada uma sexta análise. Nesta análise foram utilizado dados de 36 alunos cujas matrículas foram canceladas, com o objetivo de aplicar uma regressão linear para verificar a linearidade dos dados tomando os seguintes dados como resultados: Total de semestres, Reprovação por nota, Número de aprovação e Total de disciplinas com dispensa. Para aplicar a regressão linear abreviou-se os rótulos dos dados para: Total de Semestre = TS, Semestres Trancadas = ST, Semestre Sem Disciplinas = SD, Número de Reprovação por Nota = RN, Número de Reprovação por Falta = RF, Número de Aprovação = NA, Total de Horas Realizadas = TR, Total de Dispensas em Disciplinas = TD, Semestre de Entrada = SE, Semestre de Saída = SS, Rede de Escolar no Ensino Médio = TE.

Aplicando o Total de semestre como saída obteve-se a seguinte a regressão linear apresentada na Equação (1), com uma taxa de acurácia do resultado TS de 100%.

$$TS = 6.3 * ST - 1.6 * SD + 5.0 * RN - 1.2 * RF + 4 * NA + 6.9 * TR - 6 * TD - 1 * SE + 1 * SS - 2 * TE - 8 \quad (1)$$

Colocando reprovação por nota como saída obteve-se a Equação (2), com uma taxa de acurácia do resultado RN de 98%:

$$RN = -3.6 * ST + 1.5 * SD + 0.76 * RF + 1.3 * NA - 0.03 * TR + 1.2 * TD + 0.67 * TS - 0.4 * SE + 0.3 * SS - 0.05 * TE - 0.7 \quad (2)$$

Colocando número de aprovação como saída obteve-se a Equação (3), com uma taxa de acurácia do resultado NA de 99%.

$$NA = -0.7 * ST + 1.1 * SD + 0.1 * RN + 0.15 * RF + 0.03 * TR - 1.25 * TD - 0.5 * TS + 0.3 * SE - 0.25 * SS + 0.1 * TE - 0.1 \quad (3)$$

Colocando Total de dispensas em disciplinas como saída obteve-se a Equação (4), com uma taxa de acurácia do resultado TD de 22%.

$$TD = -0.7 * ST + 1.2 * SD + 0.05 * RN - 0.09 * RF - 0.7 * NA - 0.02 * TR - 0.6 * TS + 0.3 * SE + 0.3 * SS + 0.003 * TE \quad (4)$$

Após os resultados obtidos, que resultaram em uma tendência linear dos dados, aplicamos os dados em um algoritmo de árvore de decisão, conforme apresentado na subseção a seguir.

4.2 Árvore de Decisão

A árvore de decisão é um algoritmo de aprendizagem de máquina que, a partir de aquisição de dados e destacamento de objetivo, realiza um treinamento identificando atributos mais relevantes para atingir o objetivo determinado, simples de entender e interpretar pois as árvores podem ser visualizadas. Sua utilização foi escolhida para esse trabalho devido as suas diversas vantagens, são elas: demandar uma preparação mínima dos dados, eliminando a necessidade de normalização, criação de variáveis fictícias e exclusão de valores em branco. O custo computacional para prever dados utilizando uma árvore é logaritmico no número de pontos de dados usados para treiná-la. Além disso, essas árvores têm a capacidade de lidar tanto com dados numéricos quanto categóricos, enquanto outras técnicas muitas vezes se especializam em análise de dados com apenas um tipo de variável, elas conseguem lidar com problemas de múltiplas saídas. São modelos de caixa branca, o que significa que suas decisões são facilmente interpretáveis, fornecendo explicações lógicas para as escolhas feitas. É viável validar um modelo de árvore por meio de testes estatísticos, levando em consideração sua confiabilidade. Mesmo que as suposições do modelo sejam violadas pelo verdadeiro modelo subjacente dos dados, as árvores de decisão ainda podem apresentar um desempenho satisfatório [9]. Logo após a sexta análise, foi identificado uma linearidade de dados, tornando favorável a aplicação de um algoritmo de árvore de decisão, pela sua simplicidade e eficiência.

Para aplicação da árvore de decisão no respectivo trabalho foi implementado um algoritmo em Python com o auxílio das bibliotecas: pandas - para a leitura do arquivo .csv com os dados dos históricos dos alunos, numpy - para melhor tipagem de arquivos e scikit-learn - para a criação do modelo de árvore de decisão.

Para o treinamento do modelo, foram utilizados os seguintes dados: Semestre Sem Disciplinas, Total de Reprovação por Nota e Falta, Total de Reprovação por Nota, Média da Frequência, Total de Horas Obrigatória Realizadas, Média Geral, Total de Horas de

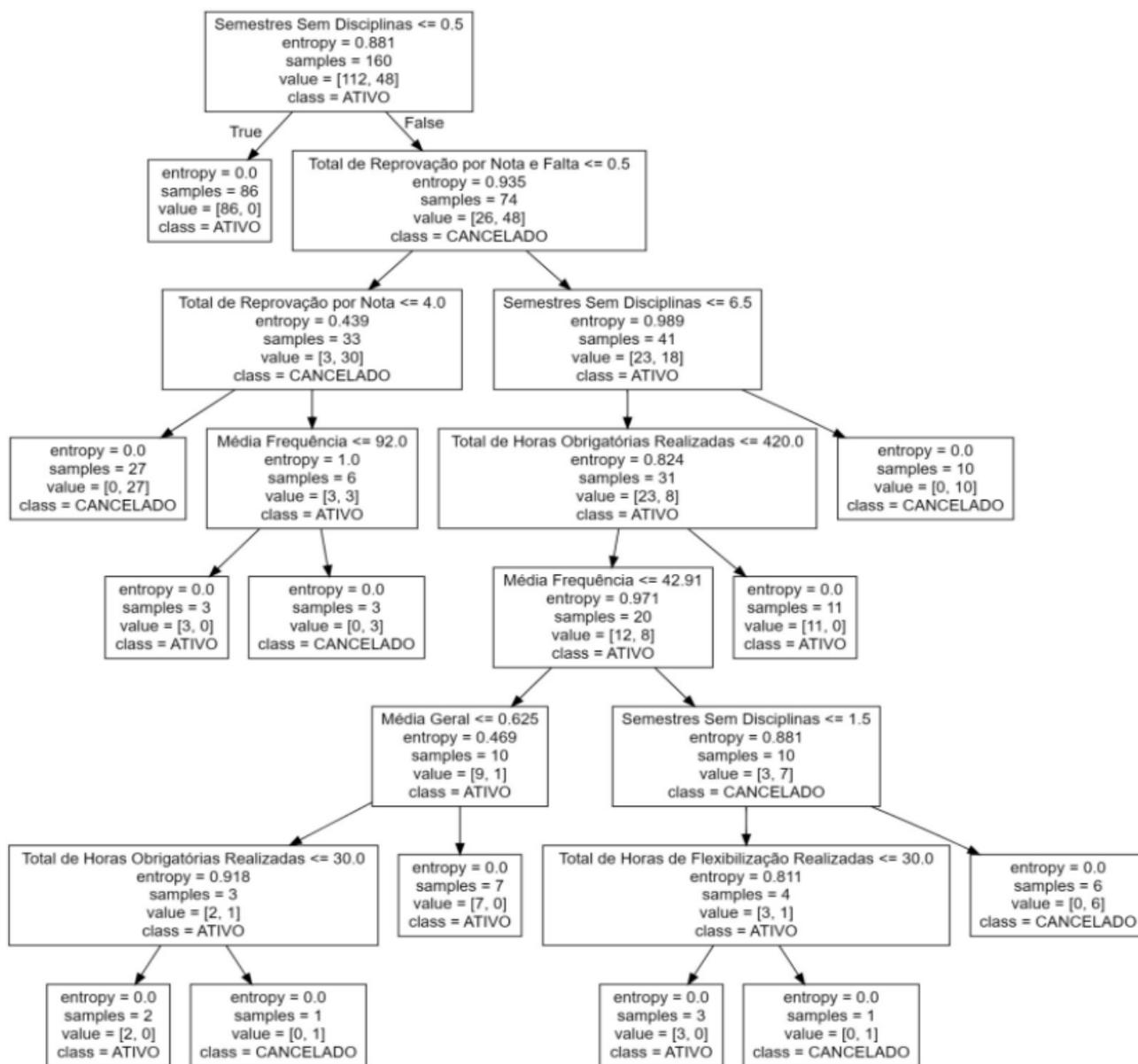


Figura 1: Árvore de decisão

Flexibilização Realizadas e Status. Ao realizar o treinamento adotou-se a coluna Status como objetivo, com Status podendo variar entre alunos ATIVOS e alunos CANCELADOS, dividimos o total de alunos (180) em dois conjuntos, sendo 90% para treinamento e 10% para teste. Após a execução do algoritmo com os dados mencionados, obteve-se a árvore de decisão apresentada na Figura 1.

Aplicando o conjunto de teste equivalente a 10% do total de alunos no modelo preditivo criado, obteve-se um total de 78% de precisão/acurácia.

4.3 Dashboard

Com o objetivo de proporcionar uma interação entre os algoritmos criados e o usuário final, optou-se pelo desenvolvimento de um *dashboard*, cujas principais telas são apresentadas nas próximas figuras.

A Figura 2 exibe a interface de login, projetada para mitigar o acesso não autorizado ao modelo preditivo. Isso é crucial, pois a exposição dos resultados por meio de características individuais pode ter um impacto emocional adverso na vida do aluno. A intenção é salvaguardar os usuários de experiências negativas ao visualizar os resultados do modelo.

A Figura 3 representa a tela liberada após a autorização prévia realizada pela tela de login. Nesta fase, presente na Figura 3, é possível fornecer os dados do discente e obter o *feedback* correspondente à sua classificação. Essa classificação pode variar entre o perfil do aluno ATIVO e o perfil do aluno CANCELADO.

As Figuras 4 e 5 ilustram o estágio subsequente à inserção e confirmação dos dados do discente. Nesse ponto, a API desenvolvida em Python é acionada, gerando o resultado proveniente do modelo preditivo mencionado anteriormente.

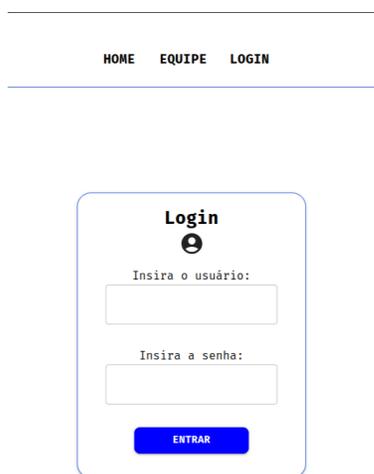


Figura 2: Tela de login.



Figura 3: Tela sucessiva ao login onde consegue acesso ao modelo preditivo.



Figura 4: Resultado positivo após preencher os dados dos alunos e consultar a API.



Figura 5: Resultado negativo após preencher os dados dos alunos e consultar a API.

5 CONSIDERAÇÕES FINAIS

O presente trabalho foi conduzido em quatro etapas distintas.

A primeira fase envolveu a extração automatizada de dados, permitindo a captura eficiente de uma grande quantidade de históricos escolares de alunos. O algoritmo criado para isso extrai automaticamente todos os dados necessários para a geração do modelo preditivo baseado em árvore de decisão que foi proposto.

A segunda fase envolveu a realização de análises de correlação, sendo que a principal preocupação era se seria possível identificar correlações significativas com a limitada quantidade de dados disponíveis. A escassez de dados representou um desafio tanto na fase de análises quanto na criação da árvore de decisão. Apesar dessa dificuldade, conseguiu-se observar uma certa linearidade nos dados,

o que motivou a implementar um modelo preditivo de árvore de decisão. Ao gerar a árvore de decisão, considerando um total de históricos analisados de 180 alunos, o modelo apresentou uma acurácia de 78%. Diante dessa limitação, é importante destacar que o modelo pode ser aprimorado, mas já oferece uma perspectiva valiosa para a intervenção precoce em alunos com possíveis indicativos de evasão escolar.

Na terceira e quarta etapa do projeto, que consistiram na implementação de uma API e de um *dashboard* interativo, a ausência de um planejamento visual adequado antes de sua implementação afetou a estética do site, que pode ser aprimorada, mas apresenta as funcionalidades inicialmente previstas.

Como uma perspectiva para trabalhos futuros, destaca-se o seguinte: dado que este trabalho explorou as relações entre os dados dos alunos que optaram por cancelar suas matrículas e aqueles que permanecem no curso, se torna necessário realizar também comparações não apenas dentro da universidade como foi o foco principal deste trabalho, mas também considerar aspectos socioambientais e regionais dos estudantes.

AGRADECIMENTOS

Os autores agradecem ao Laboratório de Iniciação Científica e Extensão da Computação, por disponibilizar sua infraestrutura e um ambiente propício para o desenvolvimento e execução das atividades de pesquisa. Ao CNPq e ao Centro Federal de Educação

Tecnológica de Minas Gerais, que através da Diretoria de Pesquisa e Pós-Graduação (DPPG) fomenta o presente projeto sob proposta de Iniciação Científica de número 10056-2022.

REFERÊNCIAS

- [1] MINISTÉRIO DA EDUCAÇÃO. Lei de diretrizes e bases da educação: Lei (9394/96). apresentação de carlos roberto jamil cury. rio de janeiro: Dp a, 2005. <https://www2.senado.leg.br/bdsf/bitstream/handle/id/70320/65.pdf>.
- [2] SENADO FEDERAL. Capítulo iii, seção i da educação, art. 205. https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.html.
- [3] P. SALDAÑA. Cerca de 4 milhões abandonaram os estudos na pandemia, diz o estudo. folha de são paulo (2021). <https://www1.folha.uol.com.br/educacao/2021/01/cerca-de-4-milhoes-abandonaram-estudos-na-pandemia-diz-pesquisa.shtml>.
- [4] Ivan Londero Hoffman et al. As informações do censo da educação superior na implementação da gestão do conhecimento organizacional sobre evasão. 2019. <https://doi.org/10.1590/0104-530X-2852-19>.
- [5] ANDIFES et al. Diplomação, retenção e evasão nos cursos de graduação em instituições de ensino superior públicas: resumo do relatório apresentado a andifes, abruem e sesu/mec pela comissão especial. avaliação: Revista da avaliação da educação superior, v.1, n. 2, 1996. [Diplomacao_Retencao_Evasao_Graduacao_em_IES_Publicas-1996.pdf](https://www.andifes.org.br/Diplomacao_Retencao_Evasao_Graduacao_em_IES_Publicas-1996.pdf)(andifes.org.br).
- [6] Luísa Behrens Palmeira Matheus Parreirasa santos. Evasão no bacharelado em ciência da computação da universidade de bráslia: análise e mineração de dados.
- [7] David Moises Barreto dos Santos Washington Pagotto Batista. Um estudo sobre a evasão no curso de engenharia de computação da uefs. <https://doi.org/10.13102/semic.v0i20.3355>.
- [8] Marques L. T. et al (2020). A evasão escolar no ensino superior: Um estudo de caso do curso de ciência da computação da ufersa / school evasion in higher education: A case study of the computer science course at ufersa. *brazilian journal of development*, 6(12), 103334–103350. <https://doi.org/10.34117/bjdv6n12-728>.
- [9] SCIKIT-LEARN. 1.10. decision trees — scikit-learn 0.22 documentation. <https://scikit-learn.org/stable/modules/tree.html>.