

15º Congresso de Inovação, Ciência e Tecnologia do IFSP - 2024

Aplicação Web para Análise do Comportamento da Evasão Escolar

GIOVANNA S. FANTACINI¹, FRANCISCO S. D. SANTOS²

¹Graduanda em Engenharia da computação, Aluno PIVICT, IFSP, Campus Birigui, giovanna.fantacini@aluno.ifsp.edu.br.

²Doutor em Agronomia, Orientador PIVICT, IFSP, Campus Birigui, fsergio.santos@ifsp.edu.br

Área de conhecimento (Tabela CNPq): 1.02.02.08-0 Análise de Dados.

RESUMO: Este estudo apresenta a implementação de uma ferramenta de análise de evasão escolar em uma Instituição de Ensino (IE) por meio da coleta, análise e interpretação de dados escolares, visando compreender e mitigar o problema, uma vez que a elevada taxa de evasão escolar representa um desafio significativo na educação, causando impactos negativos tanto para os alunos quanto para as instituições de ensino. Assim, o estudo busca criar uma ferramenta para ajudar na identificação dos fatores determinantes da evasão. O processamento dos registros escolares foi realizado utilizando técnicas de ETL (*Extract, Transform, Load*) para extrair, transformar e carregar dados relevantes, dessa forma, realizar uma análise exploratória dos dados, para identificar fatores associados à evasão. A ferramenta de análise foi disponibilizada em um site de acesso público, possibilitando que outras instituições estudem e apliquem os resultados. Essa plataforma auxilia gestores e professores na implementação de intervenções personalizadas e na avaliação contínua da eficácia dessas ações, promovendo uma educação mais inclusiva e equitativa. Os resultados demonstraram que abordagens orientadas pela ciência de dados podem efetivamente apoiar a implementação de estratégias direcionadas para a redução da evasão escolar.

PALAVRAS-CHAVE: Ciência de Dados; Análise de Dados; Evasão Escolar.

Web Application for Analyzing School Dropout Behavior

ABSTRACT: This study presents the implementation of a tool for analyzing school dropout rates in an Educational Institution (EI) through the collection, analysis, and interpretation of school data, aiming to understand and mitigate the problem, as the high dropout rate represents a significant challenge in education, causing negative impacts on both students and educational institutions. Thus, the study seeks to create a tool to help identify the determining factors of dropout. The processing of school records was carried out using ETL (*Extract, Transform, Load*) techniques to extract, transform, and load relevant data, enabling exploratory data analysis to identify factors associated with dropout. The analysis tool was made available on a publicly accessible website, allowing other institutions to study and apply the results. This platform assists managers and teachers in implementing personalized interventions and in the continuous evaluation of the effectiveness of these actions, promoting more inclusive and equitable education. The results demonstrated that data-driven approaches can effectively support the implementation of strategies aimed at reducing school dropout rates.

KEYWORDS: Data Science; Data Analysis; School Dropouts.

INTRODUÇÃO

A evasão escolar representa uma grande preocupação em todos os níveis de ensino, constituindo um desafio considerável para o sistema educacional e um obstáculo à equidade e ao acesso à educação (BATISTA; SOUZA; OLIVEIRA, 2009). Este fenômeno é influenciado por fatores socioeconômicos, familiares, pessoais e acadêmicos, que afetam a permanência dos alunos na escola (NIEROTKA; SALATA; MARTINS, 2023), (QUEIROZ, 2006). Em uma Instituição de Ensino (IE), a evasão escolar prejudica o desenvolvimento acadêmico e pessoal dos alunos, além de afetar a eficácia e a qualidade do ambiente educacional (NERI, 2015).

Neste contexto, é essencial que gestores e professores tenham acesso a informações para entender o comportamento da evasão escolar e melhorar a tomada de decisões, com o objetivo de reduzir os índices de evasão. Uma abordagem possível é a aplicação da ciência de dados, que consiste em uma variedade de tecnologias, incluindo análise de dados, mineração de dados, aprendizado de máquina e metodologias relacionadas à Ciência da Computação, Matemática e Estatística, além do conhecimento específico sobre o problema, para manipular, analisar e interpretar dados e gerar novas conclusões ou ideias (SARKER, 2021). Dessa forma, permite uma análise abrangente dos dados acadêmicos, demográficos e comportamentais dos alunos (CASTRO; FERRARI, 2016).

Compreendendo o comportamento da evasão nos cursos, gestores e professores podem implementar intervenções preventivas e direcionadas, adaptadas ao contexto específico de cada curso, para atender às necessidades individuais dos alunos e melhorar os índices de evasão escolar. Além disso, a análise de dados pode ajudar na avaliação da eficácia dessas intervenções ao longo do tempo, permitindo ajustes contínuos e refinamentos nas estratégias adotadas.

Ao empregar técnicas e ferramentas que convertem dados em informações relevantes, é possível obter uma compreensão mais ampla dos elementos que influenciam a evasão escolar nos cursos (CASTRO; FERRARI, 2016).

Portanto, este estudo busca criar uma ferramenta online e de acesso público, que consiga gerar análises gráficas dos dados, com o intuito de entender e abordar a evasão escolar em uma IE, destacando o papel crucial do uso da tecnologia da informação na promoção de uma educação mais inclusiva, equitativa e de qualidade para todos os alunos.

MATERIAIS E MÉTODOS

Para a construção do site, foi utilizada a linguagem de programação Python com a plataforma Streamlit. Esta é open-source e facilita a criação de aplicativos web interativos a partir de scripts em Python, sendo especificamente voltada para cientistas de dados e analistas. A plataforma suporta a integração com bibliotecas populares de ciência de dados e aprendizado de máquina, como pandas, NumPy e scikit-learn, possibilitando a visualização e manipulação de grandes volumes de dados em tempo real, promovendo a interação ágil e a colaboração, e permitindo que os usuários compartilhem suas aplicações de maneira eficiente através de URLs (STREAMLIT, s.d.).

Para a criação dos dados iniciais do site, foram utilizados os dados de uma IE, com as seguintes informações sobre os alunos: número de matrícula, nome do aluno, ano de ingresso, ano previsto de conclusão, deficiência, forma de ingresso, modalidade, curso, período de ingresso, etnia, percentual de progresso, sexo, tipo de escola de origem, turno, cidade e estado de origem, e datas de eventos

como matrícula, integralização, colação, defesa do TCC e conclusão do curso. Também incluíam itens pendentes para conclusão, como colação de grau, prática profissional, atividades complementares, carga horária de TCC, carga horária de disciplinas eletivas, carga horária optativa, carga horária obrigatória, ENADE e registro de TCC. O Número de matrícula e Nome do aluno não são informações necessárias para o processamento dos dados na plataforma, portanto, devem ser deletados do arquivo que será carregado.

Para a construção da análise dos dados, foi necessário inicialmente realizar o processo de *Extract, Transform and Load* (ETL). Esta é uma metodologia de processamento de dados amplamente utilizada para a integração e preparação de informações provenientes de diversas fontes. No ETL, os dados são extraídos de suas origens, transformados para atender às necessidades específicas da análise e, posteriormente, carregados em um sistema de armazenamento adequado (AGGARWAL; SINGHAL, 2022).

A partir dos dados iniciais, foi realizada a etapa de pré-processamento ou tratamento de dados, para garantir a qualidade e a integridade das informações utilizadas na análise. Esta fase envolveu a identificação e correção de inconsistências nos dados, como valores ausentes e duplicidades, bem como a normalização e padronização das variáveis relevantes. Essa etapa foi realizada utilizando a linguagem de programação Python, com a biblioteca Pandas, que proporciona ferramentas para manipulação de dados.

O pré-processamento incluiu a redução do número de colunas da base, eliminando aquelas que não continham dados válidos ou eram irrelevantes para o estudo, a padronização de termos, todos os campos de situação no curso que iniciavam com "Matrícula" foram padronizados como "Matriculado", os que tinham "Concluído" foram padronizados como "Formado", e os que iniciavam com "Cancelado" ou "Cancelamento" foram padronizados como "Evasão". Na coluna "Sexo", "M" foi substituído por "Masculino" e "F" por "Feminino".

Os dados, após processados, ficam armazenados temporariamente no site; apenas os dados iniciais são permanentes. Os dados que são inseridos para as análises são excluídos ao fim da sessão.

RESULTADOS E DISCUSSÃO

O projeto de pesquisa resultou na publicação de um site por FANTACINI (2024), como mostrado na Figura 1, que exhibe uma parte da página inicial do site com as análises gráficas da instituição de forma geral. O site é dividido em análises da IE de modo geral e em análises por curso.

As análises da IE de modo geral incluem: evasão por Ano de Ingresso, Etnia, Escola de Origem, Gênero e Cidade; análises dos cursos com maiores taxas de evasão, total de alunos matriculados por modalidade na instituição; e informações sobre o perfil dos alunos matriculados.

As análises por curso buscam, de forma mais detalhada, mostrar em qual momento os alunos desistem do curso, o perfil dos alunos no curso, os casos de atraso na formação e suas pendências. A página de cada curso possui as mesmas análises de evasão da página inicial, por Ano de Ingresso, Etnia, Escola de Origem, Gênero e Cidade, mas específicas para o curso selecionado.

A Figura 2 apresenta gráficos específicos para cada curso, ilustrando a análise da evasão por porcentagem de conclusão no curso de Engenharia da Computação da IE utilizada como exemplo no site; nela, é possível observar que de todos os alunos que ingressaram no curso 36,9% evadiram, e que, dentre esses, a grande maioria das evasões ocorre no início do curso. Além disso, mostra a situação dos alunos no curso.



Figura 1: Página Inicial do site.

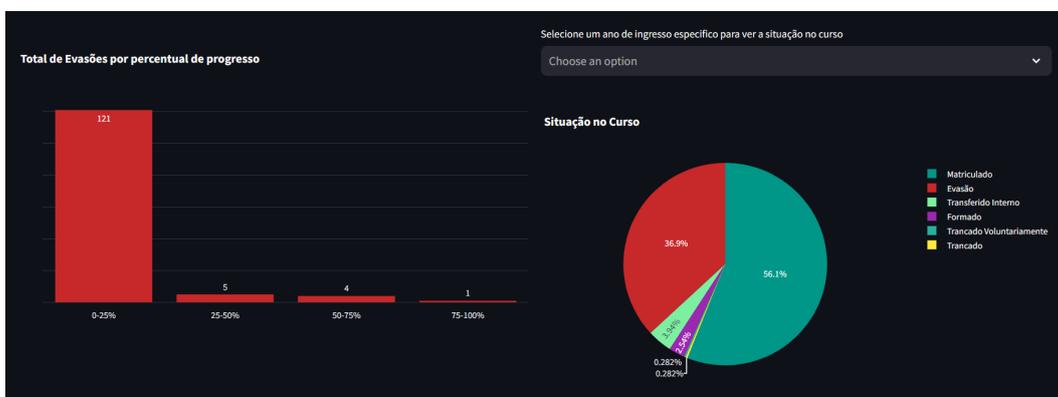


Figura 2: Análise da Evasão e Situação Acadêmica no Curso de Engenharia da Computação.

Ademais, a Figura 3 apresenta o gráfico de formação dos alunos, mostrando o total de alunos formados por ano de ingresso e comparando o número de formados dentro e fora do prazo. No exemplo apresentado, é possível notar que a maioria dos alunos se formou fora do prazo. O gráfico de barras também exibe o total de alunos matriculados por ano de ingresso, onde é possível observar que alguns alunos, que já deveriam ter se formado, ainda permanecem matriculados.



Figura 3: Análise de Conclusão no Curso de Engenharia da Computação.

A Figura 4 exibe o total de alunos por semestre em atraso e as pendências correspondentes. As informações podem ser filtradas pelo ano de previsão de conclusão dos alunos em atraso, permitindo visualizar apenas as pendências relacionadas a esse ano específico. Essa funcionalidade pode ser utilizada por gestores e professores para direcionar ações focadas nas necessidades desses alunos. Além disso, na Figura 5 há um gráfico que apresenta a taxa de evasão por forma de ingresso, acompanhada de uma

tabela com as 10 principais formas de ingresso que possuem as maiores taxas de evasão. Essa, junto com a tabela, facilita a análise do número de alunos que ingressaram e posteriormente abandonaram o curso, permitindo observar como essas saídas impactam a taxa de evasão.



Figura 4: Análise do Atraso na Formação no Curso de Engenharia da Computação.



Figura 5: Taxa de Evasão por Forma de Ingresso no Curso de Engenharia da Computação.

A página dos cursos também exibe um gráfico similar ao da Figura 1, apresentando o perfil dos alunos matriculados no curso, além de um gráfico adicional que mostra o perfil dos alunos formados. Conhecer o perfil dos alunos matriculados, como a cidade de origem, é fundamental para considerar questões importantes para o curso, como o transporte até a faculdade. Isso permite que gestores e professores desenvolvam estratégias que atendam melhor às necessidades dos estudantes, contribuindo para um ambiente acadêmico mais inclusivo e acessível.

Além dessas páginas, o site também conta com uma página de *upload* de arquivo, onde estão disponíveis todas as instruções necessárias para a inserção de um arquivo contendo os campos requeridos para a realização das análises. Após o upload, todas as análises do site são geradas com base no arquivo enviado, sobrepondo o estudo anterior.

CONCLUSÕES

Como resultado do estudo, foi publicado um site por FANTACINI (2024), o qual mostrou que a metodologia empregada, orientada à ciência de dados, pode ser eficaz na identificação dos fatores que afetam a permanência dos alunos, permitindo uma análise detalhada e visual dos dados educacionais. A ciência de dados permitiu entender o comportamento da evasão, quais alunos estão sujeitos a ela e, a partir desse cenário, colaborar com os gestores na tomada de decisão. A elaboração dessa ferramenta permite que gestores e educadores implementem estratégias de intervenção mais eficazes e conheçam melhor o perfil dos alunos, promovendo maior diversidade e inclusão. O estudo ressalta a importância da tecnologia da informação na educação e o impacto positivo da ciência de dados na criação de um ambiente educacional mais adaptado às necessidades dos alunos.

CONTRIBUIÇÕES DOS AUTORES

F.S.D.S e G.S.F. contribuíram com a concepção e escopo do estudo. G.S.F procedeu com o desenvolvimento e escreveu o trabalho.

Todos os autores contribuíram com a revisão do trabalho e aprovaram a versão submetida.

AGRADECIMENTOS

Agradecemos ao Programa Institucional Voluntário de Iniciação Científica e Tecnológica do Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (PIVICT) pelo apoio ao desenvolvimento desse projeto.

REFERÊNCIAS

- AGGARWAL, A.; SINGHAL, B. Etl, elt and reverse etl: A business case study. Second International Conference on Advanced Technologies in Intelligent Control, Environment, Computing and Communication Engineering ICATIECE, 2022. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/10046997>>. Acesso em: 19 de Set. de 2023.
- BATISTA, S. D.; SOUZA, A. M.; OLIVEIRA, J. M. S. A evasão escolar no ensino médio: um estudo de caso. *Revista Profissão Docente, UNIUBE*, Uberaba/MG, v. 19, p. 9, 2009.
- CASTRO, L. N.; FERRARI, D. G. *Introdução à Mineração de Dados: conceitos básicos, algoritmos e aplicações*. 1. ed. São Paulo: Saraiva, 2016. 40-95 p.
- FANTACINI, G. S. *Análise de Evasão Escolar*. 2024. Disponível em: <<https://apptcc.streamlit.app/>>. Acesso em: 17 de jul de 2024.
- NERI, M. *Motivos da evasão escolar*. 2015. Disponível em: <https://www.cps.fgv.br/ibrecps/rede/ finais/Etapa3-Pesq_MotivacoesEscolares_sumario_principal_anexo-Andre_FIM.pdf>.
- NIEROTKA, R. L.; SALATA, S.; MARTINS, M. K. Fatores associadas à evasão no ensino superior: Um estudo longitudinal. *Caderno de Pesquisa*, v. 53, 2023.
- QUEIROZ, L. D. Um estudo sobre a evasão escolar: para se pensar na inclusão escolar. *Revista Brasileira de Estudos Pedagógicos*, v. 64, n. 147, p. 38-69, 2006.
- SARKER, I. H. Data science and analytics: An overview from data-driven smart computing, decision-making and applications perspective. *SN Computer Science*, 2021. Disponível em: <<https://link.springer.com/article/10.1007/s42979-021-00765-8>>. Acesso em: 19 de Set. de 2023.
- STREAMLIT. s.d. Disponível em: <<https://streamlit.io/>>. Acesso em: 07 de Ago. de 2024.