

ESTUDO SOBRE TÉCNICAS COMPUTACIONAIS E ANÁLISE DOS DADOS PARA DIMINUIÇÃO DA EVASÃO ESCOLAR ¹

Gabriel dos Reis Vaichulonis², Isabela Gasparini³

¹Vinculado ao projeto “Tecnologias educacionais e inteligentes para engajamento dos estudantes e redução da evasão escolar”

² Acadêmico do Curso de Engenharia Elétrica – CCT – Bolsista PIBIC

³ Orientadora, Departamento de Ciências da Computação – CCT – isabela.gasparini@udesc.br

O presente projeto de pesquisa tem como objetivo testar e selecionar uma das técnicas, encontradas por um mapeamento sistemático, para predição da evasão universitária, assim como a análise dos dados universitários de alunos do Centro de Ciências Tecnológicas (CCT) da Universidade do Estado de Santa Catarina (UDESC). Um mapeamento sistemático é definido como a construção de um esquema de classificação e estruturação de um campo de interesse. As técnicas de predição são *scripts* que analisam um banco de dados e retornam desde qual variável contribui mais para o resultado esperado, até qual será o resultado baseado no perfil do aluno.

Dentre os métodos para a predição da evasão podemos citar *machine learning*, *data mining*, *deep learning*, sendo o foco deste trabalho o *data mining*. Data mining é um processo de identificação de padrões em conjuntos de dados usando métodos que envolvem *machine learning*, estatística e sistemas de base de dados.

Neste trabalho foi utilizado o programa Weka para iniciar a predição em bancos de dados limpos e prontos para o uso. Neste projeto foram utilizados 2 bancos de dados educacionais públicos sendo eles vindos da *Euro Students* e da *Organisation for Economic Co-operation and Development*, os quais precisam ser limpos de dados não necessários e faltantes.

Após esse primeiro passo foram recebidos dados gerais anonimizados em planilhas dos alunos que se matricularam na UDESC desde fevereiro de 2009 até novembro de 2020. Uma das planilhas é sobre perfil do aluno, nele tem-se: matrícula do aluno, curso, data de entrada, tipo de entrada, estado da matrícula, data de saída, sexo, data de nascimento, estado civil e etnia. Um segundo banco de dados contém a trajetória dos alunos, ou seja, as matérias que o aluno fez durante o curso, nele se encontram: matrícula do aluno, semestre letivo em que fez a matéria, código da disciplina, nome da disciplina feita, estado do aluno na matrícula e a média ao final do semestre na matéria.

O processo de limpeza dos dados foi realizado, assim como o ajuste dos códigos das disciplinas, pois ao longo dos 11 anos muitas matérias tiveram seus códigos alterados, então foi necessário ajustar todos os códigos de cada matéria para um único, para ser possível identificar a matéria pelo seu código. A partir dos bancos de dados resultantes das limpezas realizadas, diversas análises surgiram, como: a evasão por cada curso da UDESC (Figura 1), a evasão por semestre de entrada do aluno, a evasão pela idade do aluno (Figura 2), a evasão por matéria e a evasão pelo tipo de entrada. Os alunos que foram considerados evadidos são os que cancelaram, jubilaram, desistiram ou fizeram transferência externa. Os cursos na Figura 1 são mostrados pelos seus códigos, ou seja: Bacharelado em Engenharia de Produção (PRO-BAC); Tecnólogo em análise e desenvolvimento de sistemas (SIJ-TEC); Licenciatura em Física (LEF-LIC);

Bacharelado em Engenharia Elétrica (ELE-BAC); Bacharelado em Engenharia Mecânica (MEC-BAC); Bacharelado em Ciência da Computação (CCI-BAC); Licenciatura em Matemática (MAT-LIC); Licenciatura em Química (QUI-LIC).

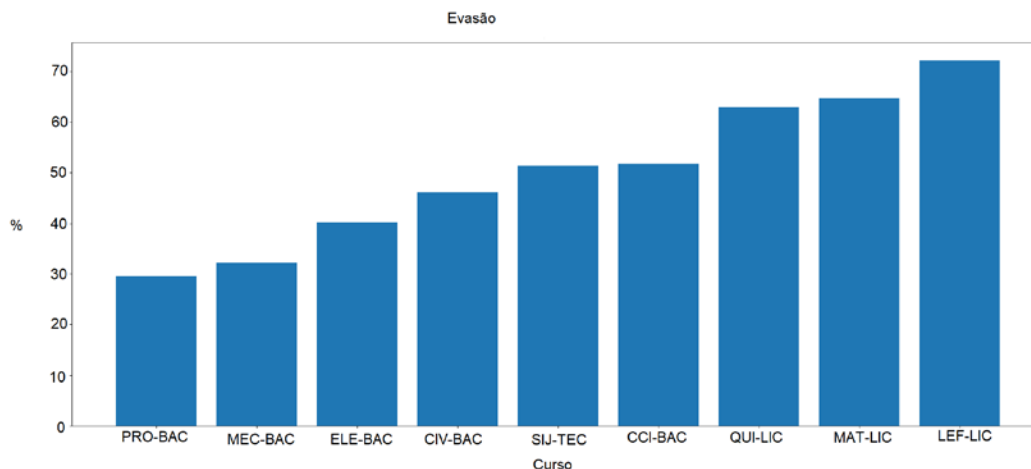


Figura 1. Evasão por cada curso da UDESC

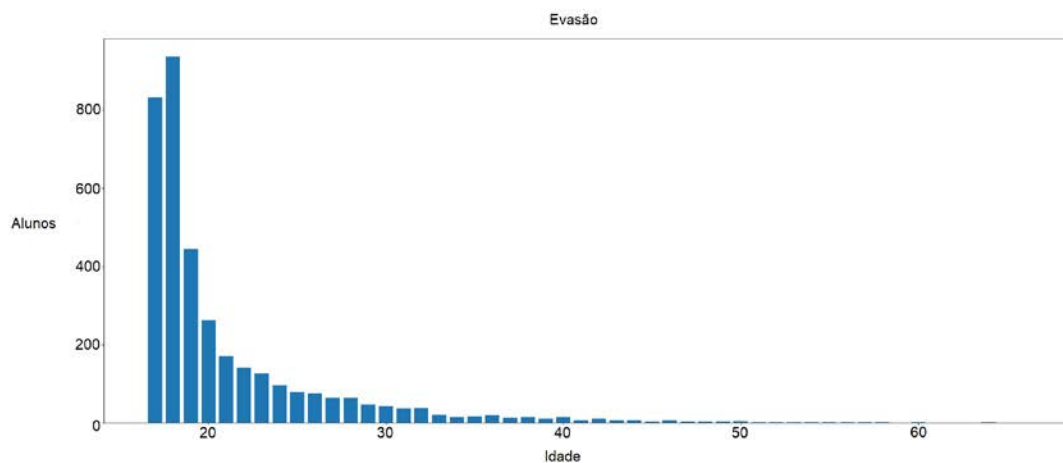


Figura 2. Evasão pela idade do aluno

Palavras-chave: Evasão. Análise. UDESC.