

# PREDIÇÃO DE ESTUDANTES EM RISCO DE RETENÇÃO E EVASÃO: UMA PROPOSTA DE ABORDAGEM EXPLORATÓRIA DOS DADOS DO IFSUL

PE01200520/008

Geison de Lima Martins (Discente - IFSul Câmpus Pelotas - Visconde da Graça - CT em Desenvolvimento de Sistemas - geison.lmartins@inf.ufpel.edu.br)  
Emanuel Marques Queiroga (Coorientador - IFSul Reitoria - DTI - emanuelqueiroga@ifsul.edu.br)  
João Ladislau Barbará Lopes (Docente Orientador - IFSul Câmpus Pelotas - Visconde da Graça - CT em Desenvolvimento de Sistemas - joaolopes@ifsul.edu.br)

INSTITUTO FEDERAL SUL-RIO-GRANDENSE (IFSUL) – CÂMPUS PELOTAS - VISCONDE DA GRAÇA

14<sup>o</sup>  
JIC  
IFSul

JORNADA DE  
INICIAÇÃO CIENTÍFICA DO  
INSTITUTO FEDERAL SUL-RIO-GRANDENSE

2021

INSTITUTO  
FEDERAL  
Sul-rio-grandense

## Introdução

Um dos atuais desafios na área da educação é a retenção e a evasão de estudantes.

No Brasil, os Institutos Federais apresentam taxas de evasão anual entre 24% e 28%, podendo atingir até 70% dos alunos matriculados até o término do curso (GUERRA *et al.*, 2019).

A detecção precoce destas situações pode ser um fator preponderante para sua reversão (QUEIROGA *et al.*, 2020).

## Objetivo

Proposição de uma abordagem exploratória dos dados de fontes institucionais do IFSul, visando gerar modelos para predição de estudantes em risco de retenção e evasão.

## Metodologia

Constituída pelas etapas de análise exploratória dos dados, preparação dos dados, geração dos modelos e avaliação dos resultados dos modelos. Foram analisados dados anonimizados de 11417 estudantes de 19 cursos técnicos e 11 cursos superiores.

## Resultados

Os resultados parciais, obtidos pela análise dos dados, possibilitaram definir quais variáveis podem ser usadas na geração dos modelos.

Nos testes iniciais com dados dos cursos técnicos (vide Tabela 1), a aplicação dos algoritmos classificadores baseados em aprendizagem de máquina obteve taxas de até 80,4% na predição de estudantes que não finalizaram o curso no tempo previsto (“d”).

Tabela 1. Matriz de confusão para predição nos cursos técnicos

		Predição		$\Sigma$
		a	d	
Atual	a	56,1%	19,6%	1134
	d	43,9%	80,4%	2813
$\Sigma$		990	2957	3947

Fonte: Elaboração própria

Já para os estudantes que concluíram o curso normalmente (“a”), o algoritmo de predição obteve uma taxa de 56,1% de acerto. Assim, totalizando uma média de 73% de precisão.

## Considerações

Um problema enfrentado na análise de dados foi a falta de preenchimento significativo em diversos campos da base. Isto gera uma menor eficiência nos modelos de predição, além de não revelar informações que poderiam ser relevantes para tratar o problema da desvinculação estudantil. Cabe salientar que a cultura de recolhimento de dados é relativamente nova no contexto educacional, sendo necessária a conscientização sobre sua importância.

Na continuidade da pesquisa pretende-se expandir a geração dos modelos aos cursos superiores e realizar a avaliação geral dos modelos de predição.

## Referências

GUERRA, L.; FERRAZ, R.; MEDEIROS, J. Evasão na educação superior de um Instituto Federal do nordeste brasileiro. *Revista Eletrônica de Educação*, v. 13, n. 2, p. 533-553, 2019.

QUEIROGA, E.; LOPES, J.; KAPPEL, K.; AGUIAR, M.; ARAÚJO, R.; MUNOZ, R.; VILLARROEL, R.; CECHINEL, C. A Learning Analytics Approach to Identify Students at Risk of Dropout: A Case Study with a Technical Distance Education Course. *Applied Sciences-Basel*, v. 10, n. 11, p. 3998, 2020.

REALIZAÇÃO  
propesp

INSTITUTO FEDERAL  
Sul-rio-grandense