

Modelos preditivos: aliados eficazes na previsão da evasão estudantil?

Jeovani Schmitt^{1,2}; Anny Key de Souza Mendonça¹; Antonio Cezar Bornia¹; Silvana Ligia Vincenzi¹; Adriano Ferreti Borgatto¹; Dalton Francisco de Andrade¹

¹ Universidade Federal de Santa Catarina – Florianópolis – SC, ² Instituto Federal Catarinense – Blumenau - SC

INTRODUÇÃO

A evasão escolar ainda é um problema mundial enfrentado pelas Instituições de Ensino (IE) [5,7]. Identificar com antecedência estudantes propensos a evadir, e planejar ações para minimizar taxas de evasão tem sido um trabalho desafiador para gestores educacionais. Apesar que geralmente as Instituições possuem muitos dados armazenados em suas bases de dados, percebe-se ainda a dificuldade de transformar esses dados em informações potencialmente úteis para auxiliar na gestão cotidiana do problema. Nesse sentido, por meio dos algoritmos de aprendizado de máquina, avanços da inteligência artificial (IA) têm ganhado destaque na previsão de eventos como no caso da evasão [1, 3, 4].

OBJETIVO

Explorar a acurácia de um algoritmo de IA para prever o risco de evasão dos estudantes de cursos superiores na modalidade presencial.

MÉTODO

A coleta dos dados ocorreu por meio da aplicação de um questionário contendo 30 perguntas que foram aplicadas para 401 estudantes no início do curso. As perguntas do questionário, denominadas variáveis previsoras, foram elaboradas a partir do modelo teórico de integração social e acadêmica de Vincent Tinto [5]. Essas perguntas investigaram a satisfação do estudante em relação à estrutura e apoio da instituição, curso e conteúdo, ambiente escolar e fatores pessoais [7]. Decorrido o prazo de integralização do curso, o status da matrícula (concluído/evadido) desses alunos foi consultado no registro acadêmico da IE, e constatou-se que 255 estudantes evadiram e 146 concluíram o curso. As 401 respostas dos estudantes para as 30 perguntas e o verdadeiro status de matrícula deles formaram uma base histórica completa submetida ao algoritmo naive bayes. Esse algoritmo divide aleatoriamente a base completa em duas: base treinamento (75%) e base previsão (25%). A base de treinamento é utilizada pelo algoritmo realizar a aprendizagem de máquina, enquanto a base previsão é destinada para testar a aprendizagem, ou seja, a classificação do status de matrícula a partir das variáveis previsoras. A aprendizagem consiste em gerar uma tabela de probabilidades para fazer novas classificações em que consta a frequência dos preditores com relação às variáveis de saída. Então, o cálculo final leva em conta a maior probabilidade para fazer a classificação.

RESULTADOS

A tabela de probabilidades gerada no aprendizado de máquina a partir dos padrões de resposta dos alunos permitiu realizar as classificações do modelo do algoritmo naive bayes. A acurácia do modelo foi 69%. Observou-se na matriz de confusão [2], que dentre os alunos classificados como evadido, 82,3% realmente não concluíram o curso, representando um resultado satisfatório.

	Status real da situação acadêmica ⁽¹⁾		Acurácia = $\frac{42+27}{100} = 69,0\%$
	Grupo 1 Evadido	Grupo 2 Concluído	
Evadido	42	9	
Concluído	22	27	
Total	64	36	N = 100

(1) situação da matrícula do estudante consultada no sistema acadêmico após o prazo de integralização do curso

DISCUSSÃO

Estudos apontam que a evasão é um fenômeno multidimensional, influenciado por fatores psicológicos, sociais, econômicos, psicopedagógicos, institucionais e didáticos.

Associadas às técnicas estatísticas que empregam modelos preditivos na avaliação do risco de evasão estudantil, as potencialidades proporcionadas pela inteligência artificial emergem como uma abordagem promissora para as Instituições de Ensino (IE). Com origem na área de *educational data mining*, esses modelos geram escores, classificações de risco ou escalas de medida, permitindo criar alertas eficazes, facilitando intervenções proativas por parte de gestores educacionais, orientando decisões e ações institucionais. A aplicação de modelos de previsão da evasão com técnicas de *machine learning* desempenham um papel importante para identificar estudantes em risco de evadir, e seus algoritmos têm mostrado bom desempenho nessa tarefa [6]. Dessa forma, encoraja-se que as Instituições de Ensino utilizem sistemas de alerta precoce (Early Warning System - EWS) como ferramentas que usam dados prontamente disponíveis em suas bases de dados para determinar quais estudantes estão mais propensos a evadir.

CONCLUSÃO

O conhecimento antecipado do risco de evasão contribui para que as IE realizem intervenções preventivas, promovendo a permanência de seus estudantes. Os algoritmos de IA não resolvem o problema da evasão, mas podem ser ferramentas potencialmente úteis e de simples implementação e aplicação cotidiana.

REFERÊNCIAS

- [1] K. S. Selim and S. S. Rezk, "On predicting school dropouts in Egypt: A machine learning approach," *Education and Information Technologies*, vol. 28, pp. 9235–9266, 2023.
- [2] I. Eegdeman, I. Cornelisz, M. Meeter, and C. van Klaveren, "Identifying false positives when targeting students at risk of dropping out," *Education Economics*, vol. 31, pp. 313-325, 2023.
- [3] J. G. C. Krueger, A. D. Britto, Jr., and J. P. Barddal, "An explainable machine learning approach for student dropout prediction," *Expert Systems with Applications*, vol. 233, p. 9, Dec 2023.
- [4] I. Eegdeman, I. Cornelisz, C. van Klaveren, and M. Meeter, "Computer or teacher: Who predicts dropout best?," *Frontiers in Education*, vol. 7, p. 10, Nov 2022.
- [5] V. Tinto, *Leaving college: Rethinking the causes and cures of student attrition*: University of Chicago press, 2012.
- [6] M. F. Tete, M. D. Sousa, T. S. de Santana, and S. Fellipe, "Predictive models for higher education dropout: A systematic literature review," *Education Policy Analysis Archives*, vol. 30, p. 24, 2022.
- [7] J. Schmitt, M. I. Fini, C. Bailer, R. Fritsch, and D. F. Andrade, "WWH-dropout scale: when, why and how to measure propensity to drop out of undergraduate courses," *Journal of Applied Research in Higher Education*, vol. 13, pp. 540-560, 2020.