

PREDIÇÃO DE EVASÃO UNIVERSITÁRIA: UM MAPEAMENTO SISTEMÁTICO

Piero Silva Salaberri, discente de pós-graduação, Universidade Federal do Pampa, Campus Bagé

Sandra Dutra Piovesan, docente, Universidade Federal do Pampa

Valesca Brasil Irala, docente, Universidade Federal do Pampa

e-mail primeiro autor- pierosalaberri.aluno@unipampa.edu.br

O trabalho identifica, por meio de um mapeamento sistemático, os algoritmos utilizados na predição de evasão em diversos contextos educativos de ensino superior, indicando quais são os atributos mais frequentes nas análises preditivas testadas, bem como se existem e como são implementadas ações, no âmbito das instituições de ensino, que utilizam as informações dos modelos preditivos propostos. A revisão da literatura foi organizada sobre as seguintes questões de pesquisa: Q1 - Quais são os algoritmos mais utilizados para a predição de evasão universitária e quais são os mais eficientes? Q2 - É possível identificar quais são as características ou atributos mais relevantes na predição de evasão? Q3 - Existem práticas, em universidades, que utilizam as informações extraídas dos modelos preditivos com intuito de reduzir os índices de evasão? Foram realizadas buscas nas bases de dados Scopus e Web of Science (WoS). A escolha por essas bases se deu por serem as principais bases de artigos científicos na área do estudo. Os critérios de inclusão utilizados foram: 1) publicações completas e de acesso livre; 2) publicações dos últimos 10 anos; 3) publicações com os itens buscados no título, resumo ou palavras-chaves; 4) publicações que comparassem algoritmos ou formas distintas de identificar evasão; e 5) estudos que utilizavam técnicas de classificação de dados. A *string* de busca contemplou a conjunção das palavras-chave: *college, university, higher education, dropout, failure, predictive, educational data mining, student e undergraduate*. Na base Web of Science, foram elencados 117 artigos e na base Scopus, 46. Usando os critérios de exclusão (Quadro 2) para aplicar refino mais apurado, restaram para análise 23 artigos da base Web of Science e 9, na Scopus. Nos trabalhos elencados, chegou-se a um pequeno grupo de algoritmos, dentre as dezenas de possibilidades existentes na literatura. Dentre as opções analisadas, pode-se verificar que os algoritmos de Árvore de Decisão, *Random Forest*, Redes Neurais e *Support Vector Machines* foram os mais consistentes ao longo dos estudos. Como foi exposto, mais de um algoritmo pode desempenhar melhor performance em um mesmo conjunto de dados, ao alterarmos o foco temporal ou de público-alvo em uma mesma pesquisa, por exemplo. Para a previsão de perfis propensos a evadir, levando em consideração dados, prioritariamente, acadêmicos, em estudos contextualizados no ensino superior, com maior carga de dados demográficos e acadêmicos, mostraram-se mais promissores os algoritmo de Árvore de Decisão C4.5 ou J-48, *Random Forest* e Redes Neurais. Quando a análise passa a ser sobre as características que tem maior predominância no processo de decisão do aluno abandonar a Universidade, pode-se identificar que não há possibilidade de se desassociar, na vida do aluno, os motivos internos e externos à instituição sobre a

decisão de permanecer ou evadir. Porém, quando analisamos os motivos que fazem os alunos ingressantes evadir, podemos pontuar que a média com que entram na instituição e como são recebidos (informações sobre as instalações físicas e sobre o próprio curso são as variáveis mais relevantes para o aluno e nelas devem ser estruturadas ações de suporte do ponto de vista da instituição de ensino. Quando os alunos já passaram pelo primeiro ano do percurso formativo, a ação deve ser de monitoramento e suporte acadêmico, pois alunos com médias baixas ou alto grau de repetência são os mais prováveis de abandonarem os estudos.

Como pôde ser visto pela leitura dos trabalhos, existem muitos exemplos de iniciativas que utilizam os dados provenientes das predições de abandono para informar, principalmente, os gestores e responsáveis por áreas dentro das universidades. Essas informações visam subsidiar esses atores para que intervenham, investigando os motivos e aplicando ações com o intuito de reter os alunos. Não foi identificado estudo que apontou desvantagem no uso das informações ou na recepção das práticas pelos alunos, professores, coordenadores ou qualquer pessoa envolvida no processo.

Como considerações finais, pode-se destacar que é possível elencar em seletos grupo de algoritmos que demonstram boa eficiência na identificação de alunos com potencial de evasão e quais são as variáveis que mais se destacam no processo de mineração dos dados. Também foram identificadas soluções de variadas abordagens para dar suporte aos alunos identificados em risco de evasão, permitindo assim a promoção da permanência do estudante.

Palavras-chave: Predição; Evasão; Ensino Superior; Mineração de dados educacionais; Estudantes.