

PREDIÇÃO DA EVASÃO EM CONTEXTOS EDUCACIONAIS: UM ESTUDO PRELIMINAR¹

Daniella Martins Vasconcellos², Êrica Peters do Carmo³, Isabela Gasparini⁴

¹ Vinculado ao projeto “Novas abordagens para o ensino de lógica de programação: estudo, análise e uso das estratégias didático-pedagógicas e dos recursos tecnológicos”

² Acadêmica do Curso de Ciência da Computação – CCT – Bolsista PIBIC/CNPq

³ Acadêmica do Curso de Ciência da Computação – CCT

⁴ Orientadora, Departamento de Ciência da Computação – CCT – isabela.gasparini@udesc.br.

A evasão é um dos principais fenômenos que atingem a educação de maneira mundial. Caracteriza-se como um problema presente nos diversos contextos educacionais. No ensino formal, o impacto da evasão pode ser observado desde à educação básica até o ensino superior. Já nas modalidades de ensino não formal, esse mesmo impacto está presente principalmente nos cursos livres à distância.

Apesar de ser um problema contemporâneo, as causas e consequências da evasão vem sendo estudadas e discutidas há décadas. Nos anos 70 e 80 os trabalhos buscavam entender quais os conjuntos de fatores que poderiam influenciar na ação do aluno de evadir. Nos anos seguintes, com os avanços da Computação, soluções atreladas à tecnologia começaram serem propostas para tentar resolver os problemas presentes na Educação. Neste cenário, diversas pesquisas foram e continuam sendo feitas com o objetivo de utilizar modelos computacionais para prever os alunos que irão evadir em determinados contextos educacionais. Possibilitando assim, que os responsáveis pelo ensino possam tomar medidas contrárias à esta ação.

A partir disto, o presente trabalho teve como objetivo a realização de um mapeamento sistemático da literatura para investigar como é realizada a predição da evasão, em quais contextos educacionais ela está presente e quais são as características e fatores associados à ela, bem como as técnicas e modelos utilizados, e os critérios para a sua avaliação. Espera-se que com os resultados do mapeamento, seja possível orientar os trabalhos futuros de pesquisa e conclusão de curso dos alunos envolvidos no projeto.

O processo metodológico do mapeamento foi feito com base nas diretrizes apresentadas por Petersen et. al. [1] e consistiu nas seguintes etapas: definição das questões de pesquisa, processo de busca (definição da *string* de pesquisa e utilização nos mecanismos de busca escolhidos), processo de seleção (de acordo com critérios de inclusão e exclusão), análise dos artigos com base na extração e categorização dos dados.

No total, foram definidas quatro questões de pesquisa: **Q1:** Quais são os modelos, técnicas e algoritmos utilizados na predição da evasão na área educacional? **Q2:** Em quais contextos educacionais ocorre essa predição? **Q3:** Quais são os elementos e características utilizados na predição? **Q4:** Quais os critérios utilizados para a avaliação dos modelos, técnicas e algoritmos de predição?

Na primeira parte do processo metodológico, foram utilizados quatro mecanismos de busca (Scopus, El Compendex, IEEE Xplore e ACM Digital Library) para encontrar artigos com a string

de busca definida, que continha as palavras-chave e suas variantes: evasão, educação, predição. Os mecanismos retornaram um total de 360 artigos.

No parte de seleção, foram adicionados os artigos que continham todos os critérios de inclusão (**CI1:** Disponível para download; **CI2:** Em inglês ou português, **CI3:** Estudo primário, **CI4:** Não duplicado), e foram excluídos todos os artigos que continham pelo menos um dos critérios de exclusão (**CE1:** Estudo sem experimento ou simulação, **CE2:** Estudo sem explicação dos resultados). Esta parte do processo metodológico resultou em 77 artigos que foram analisados e serviram de base para as respostas às questões de pesquisa descritas abaixo.

Para responder à **Q1**, os artigos foram categorizados de acordo com as técnicas e algoritmos utilizados na predição da evasão. Observou-se que os estudos dividem-se em duas abordagens. A primeira é a utilização de técnicas estatísticas, como Regressão Logística, principalmente para entender quais são as características dos alunos, definidas como features no dataset do experimento, que mais influenciam na evasão. A segunda técnica é a utilização de algoritmos de machine learning (por exemplo, *Decision Trees*, *Naive Bayes*, *Support Vector Machine*, etc) para resolver um problema de classificação binária: decidir se um aluno X é do tipo 0 (evadido) ou 1 (não evadido).

A **Q2**, relacionada aos contextos em que essa predição ocorre pode ser observada numericamente. Dos 77 estudos analisados, 55 abordavam a evasão em modalidades de ensino formal (ensino médio e ensino superior) e 21 em uma modalidade específica de ensino não formal, os cursos online abertos e massivos (MOOCs).

As análises obtidas através da **Q3**, que busca investigar as características utilizadas na predição, mostram que os fatores preditivos utilizados nos estudos variam em domínio e dependem principalmente do contexto educacional. Por exemplo, quando investigados contextos de ensino formal e presencial, as características que mais destacam-se são as demográficas (idade, gênero, grupo étnico) e sociais dos alunos, bem como as informações do seu desempenho (notas, frequência, número de créditos, matérias aprovadas e reprovadas). Já nos MOOCs, principalmente por uma questão de disponibilidade, as características utilizadas são do tipo *clickstream* (i.e. eventos que mostram a interação do aluno com o curso no ambiente virtual de ensino), por exemplo: número de vídeos assistidos, questionários respondidos, postagens nos fóruns, tempo de cada sessão.

Por fim, a **Q4** apresenta um resultado atrelado à resposta da **Q2**. Como a maior parte dos trabalhos utilizam técnicas de *machine learning*, os fatores utilizados para avaliar a predição são as métricas que avaliam o desempenhos de algoritmos de classificação. As principais são acurácia, precisão, *recall* e *f-score*.

Em conclusão, com os resultados do presente trabalho, pretende-se definir e aplicar diferentes modelos de predição em diferentes contextos educacionais atendidos pela UDESC, como por exemplo os cursos de graduação e as disciplinas ministradas com o auxílio de ambientes virtuais como o *Moodle*. Entende-se a importância da pesquisa aqui documentada para que o processo de experimentação seja feito com base em diretrizes já estabelecidas na Literatura. Além disso, faz-se possível comparar os resultados obtidos com os dos estudos analisados. Por último, ressalta-se a importância dos trabalhos que darão continuidade a este para que seja possível não só identificar as causas da evasão na UDESC, mas também, possibilitar a proposta de intervenções para mitigá-la.

Palavras-chave: evasão, predição, educação.

Referências:

[1] Petersen, K., Feldt, R., Mujtaba, S., & Mattsson, M. (2008). Systematic Mapping Studies in Software Engineering. *12th International Conference on Evaluation and Assessment in Software Engineering*. 17.