

Indicadores de Evasão em Ambientes Virtuais de Aprendizagem no contexto da Educação a Distância: Um Mapeamento Sistemático

Patricia Mariotto Mozzaquatro Chicon, UNICRUZ/UNIJUI, patriciamozzaquatro@gmail.com
Leo Natan Paschoal, ICMC – USP, paschoalln@usp.br
Fabricia Carneiro Roos Frantz, UNIJUI, frfrantz@unijui.edu.br

Resumo: *Um dos desafios das instituições de ensino é reduzir os altos índices de evasão em cursos na modalidade a distância. Este artigo tem como objetivo identificar na literatura os trabalhos de pesquisa sobre indicadores de evasão no contexto da educação a distância e a relação destes indicadores com recursos e dados de ambientes virtuais de aprendizagem. Um mapeamento sistemático foi conduzido, no qual foram examinados 22 estudos primários. Os resultados obtidos revelam os principais indicadores de evasão no contexto da educação a distância, são eles: indicador comportamental, desempenho e demográfico. Os dados coletados pelos ambientes virtuais de aprendizagem, utilizados pelos indicadores, provêm de recursos como envio de tarefas, participação em fórum, interações em chat, respostas e interações em questionários, dentre outros.*

Palavras-chave: *Ambientes Virtuais de Aprendizagem, Educação a Distância, Indicadores de Evasão.*

Dropout indicators in Virtual Learning Environments in a Distance Education context: A Systematic Mapping

Abstract: *One of the challenges facing educational institutions is to reduce the high dropout rates in distance learning courses. This article aims to identify in the literature the research studies about dropout indicators in the context of distance education and the relationship of these indicators with resources and data from virtual learning environments. We conducted a systematic mapping and identified 22 primary studies on the topic. The results obtained reveal the main evasion indicators in the context of distance education such as behavioral, performance, and demographic indicators. The data collected by the virtual learning environments, used by the indicators, come from resources such as task sending, participation in forums, chat interactions, questionnaires responses, among others.*

Keywords: *Learning Management System, Distance Education, Dropout factors.*

1. Introdução

A oferta de cursos na modalidade de ensino a distância (EaD) vem crescendo nos últimos anos. De acordo com os dados da Associação Brasileira de Educação a Distância (ABED, 2019), no seu censo de 2018, a modalidade de ensino a distância em 2016 tinha 3.734.887 alunos, em 2017 um total de 7.773.828 alunos e em 2018 um total de 9.374.647 alunos. Com o crescimento desta modalidade de ensino surgiram novos desafios e oportunidades pedagógicas e computacionais, relacionadas a um melhor aproveitamento dos recursos disponíveis (QUEIROGA; CECHINEL; ARAÚJO, 2015).

Um dos principais desafios das instituições de ensino é a redução dos altos índices de evasão em seus cursos ministrados na modalidade EaD. O ensino nesta modalidade é mediado por tecnologias em que alunos e professores utilizam Ambientes Virtuais de Aprendizagem (AVAs). Por meio dessas ferramentas ocorrem as interações e o registro das ações que

são executadas pelos usuários. Os dados destes registros podem ser usados para identificar o comportamento dos alunos e apoiar a equipe de EAD a mitigar problemas que emergem no contexto dessa modalidade de ensino, tais como o problema da evasão. Dessa forma, a predição dos alunos com risco de evasão e estabelecimento de mecanismos e metodologias de apoio que possam mitigar a evasão tornou-se um dos principais tópicos de investigação na área de análise de dados educacionais. (BRITO *et al.*, 2015; DIGIAMPIETRI; NAKANO; LAURETTO, 2016). O termo *evasão* possibilita diversas interpretações e já é utilizado em contextos distintos com significados diversos. Esta pesquisa considera a definição de Chen and Zhang (2017) e Balakrishnan (2013), na qual evasão é quando o aluno permanece sem acessar e interagir com o AVA.

Identificar alunos com riscos de evasão é considerado importante, pois muito da credibilidade de uma instituição de ensino depende da sua capacidade de reter alunos (MORAIS, 2018). Afinal, se a evasão de alunos é algo comum, isso pode sinalizar que há algo de errado com a instituição, comprometendo sua credibilidade e a promoção de uma imagem negativa quanto a curso e instituição. Partindo-se deste pressuposto, torna-se necessário reconhecer os estudantes com risco de evasão tão cedo quanto possível, permitindo que ações proativas possam ser tomadas no sentido de evitar insucessos acadêmicos. Além disso, o diagnóstico precoce de alunos com risco de evadir pode ser útil para profissionais que atuam diretamente no processo de ensino-aprendizagem, tais como professores e tutores, pois assim, o aprendizado do aluno pode ser direcionado, o ajuste de metodologias que não estão funcionando, ou até mesmo fornecer mecanismos de apoio ao ensino que possam ser mais adequados ao perfil do aluno (MACFADYEN; DAWSON, 2010).

Embora a literatura apresente um número considerável de estudos que abordam a detecção de evasão na EaD, ainda existem vários desafios a serem tratados, relacionados aos indicadores, recursos e dados armazenados pelos AVAS. Há diferentes tipos de interações que ocorrem nos ambientes, tais como participação em fóruns, acesso a tarefas, envio de mensagem, dentre outros. Porém, ainda não existe um arcabouço sólido disponível para decidir quais interações específicas dos AVAs devem ser analisados (AGUDO-PEREGRINA *et al.*, 2014). Há a necessidade de incluir outros indicadores, não citados nas pesquisas, como variáveis que afetam a aprendizagem, apresentadas por Kizilcec, Pérez-Sanagustín and Maldonado (2017) e as dificuldades no processo de seleção de dados dos AVAs que influenciam na identificação do risco de evasão em cursos na modalidade EaD. (KIKI, 2018).

Frente a este cenário, este artigo tem como objetivo mapear os principais indicadores de evasão no contexto da EaD associados a recursos e dados de AVAs, a fim de elencar os indicadores que perfazem os AVAS e que podem ser considerados durante a predição de evasão em cursos EAD. O restante deste artigo está estruturado da seguinte forma: a Seção 2 apresenta os trabalhos de revisão sistemática que tratam de indicadores de evasão em AVAs no contexto da EaD. Na Seção 3, os materiais e métodos são apresentados. Na Seção 4, os resultados são analisados e discutidos. Por fim, na Seção 5, apresenta-se as principais conclusões.

2. Trabalhos relacionados

Alguns esforços foram realizados na tentativa de caracterizar as infraestruturas e metodologias que estão sendo investigadas e estabelecidas para prever a evasão. Esses esforços foram socializados para a comunidade por meio de artigos científicos de revisão de literatura. Nesta seção, foram reunidas pesquisas que se relacionam com a proposta deste estudo

pela temática e abordagem explorada.

O estudo de Khanna, Singh and Alam (2016) apresenta uma revisão sistemática sobre Mineração de dados educacional¹ (MDE). O estudo enfatiza a aplicação da MDE para prever o desempenho acadêmico dos estudantes e detectar o comportamento do aluno. Quanto aos fatores que afetam o desempenho e a aprendizagem, o estudo relata que existem diversos fatores, tais como: percepção dos alunos em relação aos conteúdos abordados, história pessoal do aluno, auto eficácia ou metacognição, frequência às aulas, idade dos alunos, status socioeconômico, dentre outros.

Tamada, Netto and Lima (2019) realizaram uma revisão sistemática dos trabalhos que utilizam aprendizado de máquina para reduzir altas taxas de evasão em ambientes virtuais de aprendizagem. A revisão relata as técnicas de aprendizado de máquina utilizadas e identifica soluções propostas para reduzir a evasão no ensino à distância. Os autores constataram que várias soluções propostas ainda não foram praticamente implementadas, mas que a maioria delas utiliza aprendizado supervisionado para a previsão do abandono escolar. Logistic Regression e Support Vector Machine são as técnicas comumente usadas em trabalhos que apresentaram alta precisão na previsão de evasão de alunos no AVA.

Marques *et al.* (2019) apresentaram um MS sobre evasão escolar, em que buscaram identificar tecnologias de MDE e possíveis causas da evasão escolar. Nas pesquisas analisadas, os autores observaram que a ferramenta Weka² é a mais utilizada para auxiliar a desvendar as causas da evasão escolar. Entre as técnicas, destacou-se a utilização da classificação, sendo que os algoritmos de classificação foram: naive bayes, support-vector machine, network of radial basis function, multilayer perceptron, k-nearest neighbours, Jrip, OneR, J48, PART e AdaBoostM1. Ainda, o mapeamento identificou que os principais trabalhos da área se concentraram em estudar fatores/causas relacionados às características individuais do aluno, como desinteresse e insatisfação com a escola ou universidade, descoberta de novos interesses, baixo desempenho, constantes reprovações, baixa frequência nas aulas, falta de habilidade de estudo, entre outros.

Lima and Roberta (2020) desenvolveram um estudo para identificar fatores que causam abandono escolar em instituições de ensino superior no Brasil. Foi realizada uma análise correlacional (Pearson e Spearman) dos fatores educacionais e econômicos presentes no Censo Escolar. Foram coletados dados de duas bases: base de Dados do Censo da Educação Superior e a dos Indicadores de Fluxo da Educação Superior. Foi realizada uma análise descritiva da evasão escolar no conjunto de dados resultantes. Ainda, o estudo propôs modelos de predição da evasão escolar levando em consideração os fatores educacionais e econômicos baseado na análise correlacional. Como resultado, os fatores relacionados à evasão escolar foram: alimentação, permanência, material didático e transporte. Como técnicas foi observado que a regressão robusta e a regressão linear apresentam menores erros na predição.

Frente ao que a comunidade tem investigado sobre evasão de alunos, não foi possível observar esforços associados com o reconhecimento e rastreamento de ações feitas pelos alunos em ambientes amplamente utilizados para a educação a distância, que possam indicar e caracterizar uma possível evasão. Portanto, pode-se assumir que até o momento não há um reconhecimento sistemático sobre elementos indicadores de evasão que se fazem presentes nos AVAs. O estudo de Tamada, Netto and Lima (2019) concentrou-se em tec-

¹Uma abordagem indutiva que cria modelos para descobrir automaticamente informações ocultas presentes nos dados dos alunos que podem ser utilizadas na melhoria da aprendizagem (ROMERO; VENTURA; GARCÍA, 2008).

²Mais informações disponíveis em: <<https://www.cs.waikato.ac.nz/ml/weka/>>.

nologias/ferramentas que utilizam aprendizado de máquina, não em indicadores de evasão. Já os estudos de Khanna, Singh and Alam (2016), Marques *et al.* (2019) e Lima and Roberta (2020) mapearam fatores que causam abandono escolar, porém, não trataram dados disponíveis em AVAs. Os mapeamentos feitos até aqui não direcionaram a atenção ao que existe nos AVAs, que são os mecanismos que promovem o EaD. Sabe-se que alguns modelos de predição têm usado somente dados dos AVAs como em Balakrishnan (2013) e Brito *et al.* (2015), porém, não se têm conhecimento sobre o vínculo que existe entre os indicadores de evasão com os recursos que estão disponíveis para aluno e professor no AVA, precisa-se ter um entendimento sobre a ligação entre os indicadores de evasão com os dados dos estudantes disponíveis no AVA e, por fim, obter informações sobre elementos e fatores que indicam evidências de uma possível evasão considerando a audiência da modalidade EAD.

3. Materiais e métodos

A proposta desta pesquisa é prover uma visão geral sobre os indicadores de evasão em cursos na modalidade EAD e a relação desses indicadores com recursos e dados que persistem nos AVAs. Para tanto, o estudo segue procedimentos descritos por Petersen *et al.* (2008) no sentido de identificar os estudos primários de modo sistematizado. Essa identificação ocorre por meio de algumas etapas, a saber: (i) definição de questões de pesquisa; (ii) busca de estudos primários; (iii) estabelecimento de critérios de inclusão e exclusão; (iv) leitura dos estudos primários; (v) extração de dados e síntese dos resultados. Cada uma dessas etapas é apresentada ao longo desta seção.

3.1. Definição de questões de pesquisa

Para que o objetivo do estudo seja atingido, questões de pesquisa foram determinadas. Em conjunto com cada questão, alguns fatores que motivaram a sua definição foram relacionados. A Tabela 1 apresenta as questões de pesquisa que norteiam o estudo, assim como os motivos pelos quais cada questão foi levantada.

Tabela 1. Questões de pesquisa

Questões de pesquisa (QP)	Motivação (M)
QP1. Quais são os indicadores de evasão no contexto da EAD?	M1. Oferecer uma visão para a comunidade sobre elementos e fatores que podem indicar que um aluno de EAD pode evadir de um curso ou disciplina.
QP2. Qual a relação dos indicadores de evasão com os recursos do AVA?	M2. Compreender o vínculo que existe entre os indicadores de evasão com os recursos que estão disponíveis para aluno e professor no AVA.
QP3. Qual a relação dos indicadores de evasão com os dados que persistem no AVA?	M3. Prover um entendimento sobre a ligação entre os indicadores de evasão com os dados dos estudantes disponíveis no AVA.
QP4. Qual a relação dos indicadores de evasão com o público-alvo?	M4. Obter informações sobre elementos e fatores que podem indicar que um aluno de EAD pode evadir considerando a audiência da modalidade EAD.
QP5. Os indicadores são usados em algum modelo de predição de evasão?	M5. Reconhecer soluções para mitigar a evasão que são relacionadas com os indicadores de evasão.

3.2. Busca de estudos primários

Após a definição das questões de pesquisa, algumas bases de dados que indexam estudos sobre a temática foram selecionadas: ACM Digital Library³, Engineering Village⁴, IEEE Xplore Digital Library⁵, ScienceDirect⁶, Scopus⁷ e Web of Science⁸.

³Mais informações disponíveis em: <<https://dl.acm.org/>>.

⁴Mais informações disponíveis em: <<https://www.engineeringvillage.com/search/expert.url>>.

⁵Mais informações disponíveis em: <<https://ieeexplore.ieee.org/search/advanced/command>>.

⁶Mais informações disponíveis em: <<https://www.sciencedirect.com/>>.

⁷Mais informações disponíveis em: <<https://www.scopus.com/search/form.uri>>.

⁸Mais informações disponíveis em: <<https://apps.webofknowledge.com/>>.

Ainda nessa etapa, uma *string* de busca foi estabelecida, de modo a contemplar termos associados à “indicadores”, “evasão” e “AVAs”. Assim, a *string* usada nesses mecanismos foi:

```
((`EVASION` OR `ABANDONMENT` OR `AVOIDANCE` OR `DROPOUT`) AND
(`VARIABLE` OR `FACTOR` OR `ATTRIBUTE` OR `INDICATORS`) AND
(`VIRTUAL LEARNING ENVIRONMENT` OR `VLE` OR `VIRTUAL TRAINING ENVIRONMENTS` OR
`LEARNING MANAGEMENT SYSTEM` OR `LMS` OR `LEARNING AND CONTENT MANAGEMENT
SYSTEM` OR `LCMS` OR `CONTENT MANAGEMENT SYSTEM` OR `CMS`))
```

Vale salientar que foi necessário adequá-la aos diferentes mecanismos de buscas selecionados e, portanto, variações dessa *string* de busca foram feitas. As variações podem ser consultadas em: <<https://bit.ly/3oMcagJ>>.

3.3. Estabelecimento de critérios de inclusão e exclusão

Uma das etapas de um MS é a definição de critérios para serem usados na seleção de estudos primários. Os estudos primários que satisfazem os critérios de inclusão são incluídos e aqueles que não satisfazem são excluídos. A Tabela 2 apresenta os critérios estabelecidos e usados neste estudo.

Tabela 2. Critério de seleção

Critérios de inclusão (CI)	Critérios de exclusão (CE)
CI1: O estudo primário apresenta uma discussão sobre os indicadores de evasão a partir de uma perspectiva de informações que persistem no contexto dos AVAs.	CE1: O estudo é um relatório técnico, um documento que está disponível em formato de resumo, uma chamada de trabalhos ou de uma conferência, um estudo secundário ou terciário.
CI2: O estudo primário está disponível em inglês ou português.	CE2: O estudo primário não atende aos critérios de inclusão.

3.4. Leitura dos estudos primários e snowballing

A partir das definições de *string*, bases de dados e critérios de seleção, as buscas foram realizadas. Na sequência, com a intenção de reconhecer e eliminar estudos duplicados, os arquivos *bibtex* recuperados nas bases de dados foram importados na ferramenta Parsifal⁹. Estudos equivalentes foram removidos e a primeira fase de leitura teve início (*i.e.*, fase de seleção). Na fase de seleção, os títulos, resumos e palavras-chave dos estudos primários foram lidos e confrontados com os critérios de seleção. Assim, estudos que não se adequaram ao propósito deste mapeamento foram descartados. Ao final da fase de seleção, os estudos incluídos foram lidos na íntegra e confrontados novamente com os critérios de seleção. Com base nisso, obteve-se o conjunto de estudos primários que compõe este estudo secundário.

Visando complementar o conjunto selecionado, a técnica de *snowballing* foi utilizada. Nesse sentido, os pesquisadores responsáveis por este estudo analisaram as referências bibliográficas de cada estudo primário selecionado e um novo processo de análise foi feito. A análise envolveu a leitura completa das referências, considerando os critérios de seleção. Por meio dessa análise, o conjunto final de estudos primários foi definido.

3.5. Extração de dados e síntese dos resultados

A última etapa do mapeamento consiste na extração de dados, no qual os dados e informações disponíveis nos estudos primários selecionados são extraídas com o propósito de responder às questões de pesquisas, assim como, é feita a sumarização dos resultados.

⁹Mais informações disponíveis em: <<https://parsif.al/>>

Para subsidiar a extração de dados, um formulário de apoio foi definido no Google Docs. A partir dessa coleta de dados, os dados foram organizados em tabelas e gráficos no sentido de caracterizar os resultados obtidos. A Seção 4 apresentará a síntese desses resultados.

3.6. Execução do mapeamento sistemático

Considerando o processo de condução, por meio das bases de dados foi possível identificar 101 estudos na busca inicial. Desse total, 15 estudos foram incluídos ao se fazer a leitura de título, resumo e palavras-chave. Os 15 estudos restantes foram lidos na íntegra e 10 deles foram selecionados por extração de dados. A Tabela 3 contém uma síntese da execução. Vale salientar que antes da extração de dados, os estudos que foram referenciados pelos 10 estudos primários selecionados também foram analisados, usando *snowballing* e, a partir disso foram incluídos 12 novos estudos. Portanto, 22 estudos constituíram a amostragem de trabalhos que discute sobre a temática abordada neste trabalho.

Tabela 3. Resumo da condução do mapeamento

	Fase de seleção		Fase de extração	
	Incluídos	Excluídos	Incluídos	Excluídos
ACM Digital Library	0	1	0	0
Engineering Village	0	12	0	0
IEEE Xplore Digital Library	0	0	0	0
ScienceDirect	1	3	0	1
Scopus	10	62	8	2
Web of Science	4	8	2	2

4. Resultados

Esta seção foi estruturada de modo a responder cada QP deste mapeamento sistemático. As respostas foram organizadas a partir de dados coletados nos estudos que são listados na Tabela 4¹⁰. Nesta tabela, foi incluído um identificador (ID) para cada estudo primário, que será utilizado para apontar a origem de cada resultado apresentado.

Quais são os indicadores de evasão no contexto da EAD?

Para responder esta questão de pesquisa, os estudos foram analisados individualmente e classificados de acordo com o indicador de evasão que se encontrou. Todos os indicadores de evasão identificados nos estudos primários estão listados na Tabela 5. A primeira coluna da tabela relaciona os indicadores de evasão, a segunda coluna lista os estudos primários que tratam cada um deles. Destacam-se os indicadores comportamentais¹¹ que foram utilizadas em 17 estudos primários, em seguida o indicador de desempenho¹² empregado em 15 estudos primários.

Qual a relação dos indicadores de evasão com os recursos do AVA?

Cada indicador de evasão identificado na QP1 foi analisado e então relacionado aos recursos do AVA, como pode ser observado na Tabela 6, primeira e segunda colunas. Posteriormente, a terceira coluna lista os estudos primários correspondentes a cada categoria. Salienta-se que a maioria dos trabalhos analisados abordam o indicador comportamental. Observou-se

¹⁰As referências dos estudos estão disponíveis em: <<https://bit.ly/322p6oT>>.

¹¹Os indicadores comportamentais são relacionados a fatores relativos à aprendizagem ativa, motivação e envolvimento dos alunos, como acesso ao curso, aos materiais lidos e sua evolução com a programação do curso (MAZZA; DIMITROVA, 2007).

¹²Os indicadores de desempenho estão relacionados ao desempenho geral dos alunos, com base em seu desempenho nas atividades (questionário, tarefas, fóruns de discussão, dentre outros) (MAZZA; DIMITROVA, 2007).

¹³Os indicadores demográficos são relacionados a fatores relativos a dados históricos como sexo, idade, estado civil, ano de graduação, status anterior de graduação, localização, profissional (HLIOUI; ALOUI; GARGOURI, 2020).

Tabela 4. Lista completa dos estudos primários

Título do estudo	ID do estudo
A report-type plugin to indicate dropout risk in the Virtual Learning Environment Moodle	[E1]
An Infographics-based Tool for Monitoring Dropout Risk on Distance Learning in Higher Education	[E2]
Classification and predictive analysis of educational data to improve the quality of distance learning courses	[E3]
Co-embeddings for Student Modeling in Virtual Learning Environments	[E4]
CourseVis: A graphical student monitoring tool for supporting instructors in web-based distance courses	[E5]
Efficient computation of key performance indicators in a distance learning university	[E6]
Identificação de Perfis de Evasão e Mau Desempenho para Geração de Alertas num Contexto de Educação a Distância	[E7]
Information management processes for extraction of student dropout indicators in courses in distance mode	[E8]
Learning Analytics in Practice: Providing E-Learning Researchers and Practitioners with Activity Data	[E9]
Learning to Identify At-Risk Students in Distance Education Using Interaction Counts	[E10]
Multi-agent System Based on Fuzzy Logic for Elearning Collaborative System	[E11]
OULAD MOOC Dropout and Result Prediction using Ensemble, Deep Learning and Regression Techniques	[E12]
Pedagogical monitoring as a tool to reduce dropout in distance learning in family health	[E13]
Predicting Students Success in Blended Learning—Evaluating Different Interactions Inside Learning Management Systems	[E14]
Predicting students' final performance from participation in on-line discussion forums	[E15]
Students' Success Predictive Models Based on Selected Input Parameters Set	[E16]
Um estudo do uso de contagem de interações semanais para predição precoce de evasão em educação a distância	[E17]
Um modelo preditivo para diagnóstico de evasão baseado nas interações de alunos em fóruns de discussão	[E18]
Uma Abordagem Genérica de Identificação Precoce de Estudantes com Risco de Evasão em um AVA utilizando Técnicas de Mineração de Dados	[E19]
Understanding Learner Engagement in a Virtual Learning Environment	[E20]
Using data mining as a strategy for assessing asynchronous discussion forums	[E21]
What can completion time of quizzes tell us about students' motivations and learning strategies?	[E22]

Tabela 5. Indicadores de evasão no contexto da EAD

Indicador	ID do estudo
Comportamental	[E1], [E2], [E3], [E5], [E7], [E8], [E9], [E10], [E11], [E12], [E14], [E15], [E16], [E17], [E18], [E20], [E21]
Demográfico ¹³	[E6], [E7], [E13], [E16], [E20]
Desempenho	[E1], [E2], [E4], [E5], [E9], [E12], [E14], [E15], [E16], [E17], [E18], [E19], [E20], [E21], [E22]

que o comportamento do aluno pode ser associado a interações em chat, conteúdo, fórum, mensagem, questionário, tarefa e videoconferência em AVAS e que o desempenho do aluno poderá ser medido pela participação em fórum, envio de questionário e envio de tarefas já relacionados ao indicador comportamental. Em contrapartida, o indicador demográfico tem relação apenas com o perfil do estudante. O estudo primário [E8] não identifica o recurso do AVA de onde provêm os dados, portanto este estudo não foi inserido na tabela.

Qual a relação dos indicadores de evasão com os dados que persistem no AVA?

Para responder sobre a relação dos indicadores de evasão com os dados que persistem no AVA, um esquema de classificação foi organizado, neste caso por indicador–dado. Pode-se observar que no indicador demográfico foram utilizados somente dados históricos (*e.g.*, sexo, idade, estado civil, ano de graduação, status anterior de graduação, localização, profissional) que são encontrados no recurso perfil do AVA. O indicador comportamental

Tabela 6. Relação dos indicadores de evasão com os recursos do AVA

Indicador	Recurso	ID do estudo
Demográfico	Perfil	[E13], [E20], [E16], [E6], [E7]
	Questionário	[E2], [E3], [E7], [E10],[E14],[E20]
	Fórum	[E1], [E2], [E3], [E5], [E7], [E9], [E10], [E11], [E12], [E14], [E15], [E17],[E18], [E20], [E21]
Comportamental	Mensagem	[E1], [E7], [E9], [E10], [E11], [E12], [E14], [E17]
	Tarefa	[E2], [E3], [E7], [E10], [E16]
	Conteúdo	[E2],[E10], [E12], [E17]
	Videoconferência	[E5]
	Chat	[E5]
Desempenho	Questionário	[E1], [E2], [E5], [E12], [E17], [E20], [E22]
	Fórum	[E9], [E15], [E18], [E21]
	Tarefa	[E2], [E4], [E5], [E12], [E14], [E16], [E17], [E19]], [E20]

relaciona-se às interações realizadas em recursos e atividades do AVA, tais como *logs*, recursos acessados e atividades acessadas. O indicador de desempenho é obtido por meio de notas em atividades, participação em fórum e questionário do AVA. Constatou-se que os dados provindos de questionários, fóruns e atividades podem ser utilizados como medida dos indicadores comportamental e de desempenho, conforme a Tabela 7.

Tabela 7. Relação dos indicadores com os dados do AVA

Indicador	Dado	ID do estudo
Comportamental	Interações	[E1], [E2], [E3], [E5], [E7], [E8], [E9], [E10], [E11], [E12], [E14], [E15], [E16], [E17], [E18], [E20], [E21]
Demográfico	Dados históricos	[E6], [E7], [E13], [E16], [E20]
Desempenho	Nota	[E1], [E2], [E4], [E5], [E9], [E12], [E14], [E15], [E16], [E17], [E18], [E19], [E20], [E21], [E22]

Qual a relação dos indicadores de evasão com o público-alvo?

A relação dos indicadores de evasão com o público alvo é ilustrada na Tabela 8. Para identificar o público-alvo, foi considerado o nível de escolaridade. Constatou-se que a maioria dos estudos foram aplicados no contexto de cursos de graduação. Observou-se que o indicador demográfico só é considerado em cursos de graduação, já os indicadores comportamental e de desempenho são considerados em cursos de graduação e técnico (médio). As pesquisas de [E1], [E3] e [E20] não identificaram o perfil do público alvo, assim, não foram incluídas na Tabela 8.

Tabela 8. Relação dos indicadores de evasão com o público-alvo

Indicador	Nível educacional	ID do estudo
Comportamental	Graduação	[E11],[E14], [E15], [E16]
Comportamental e desempenho		[E2], [E4], [E5], [E7], [E9], [E10], [E11], [E12], [E16], [E21]
Demográfico		[E6], [E13]
Desempenho		[E19], [E22]
Comportamental	Técnico	[E8]
Comportamental e desempenho		[E17], [E18]

Os indicadores são usados em algum modelo de predição de evasão?

Conforme mencionado no decorrer deste trabalho, modelos têm sido construídos para prever alunos com risco de evasão. Considerando a temática deste estudo, foi investigado se no contexto dos estudos primários selecionados, modelos de predição de evasão foram estabelecidos. Constatou-se que 16 dos estudos primários ([E2], [E3], [E4], [E7], [E9], [E10],

[E11], [E12], [E14], [E15], [E16], [E17], [E18], [E19], [E20], [E21]) utilizaram os indicadores de evasão em modelos de predição. Seis (6) estudos primários ([E1], [E5], [E6], [E8], [E13], [E22]) abordaram os indicadores, mas não mencionaram sobre o uso dos mesmos em modelo de predição de evasão.

5. Conclusões

Neste trabalho foi apresentado o planejamento, a condução e os resultados de um mapeamento sistemático sobre o problema de evasão no contexto da EaD, em que se buscou elencar indicadores de evasão no contexto da EAD e sua relação com recursos e dados de AVAs. A partir dos resultados, constatou-se que os indicadores de evasão abordados nas pesquisas foram o indicador comportamental, indicador de desempenho e o indicador demográfico. Os dados coletados pelos AVAs, utilizados pelos indicadores, provêm de recursos como participação em fórum, envio de tarefas, interações em chat, respostas em questionários, dentre outros. Constatou-se ainda que a grande maioria dos estudos utilizaram os indicadores citados em modelos de predição de evasão e que a maioria dos estudos foram aplicados no contexto da graduação.

Espera-se que este estudo de mapeamento não sirva apenas para destacar os principais indicadores de evasão e sua relação com dados e recursos de AVAs, mas sirva para atrair pesquisadores e profissionais para descobrir um corpo de conhecimento que identifica principais indicadores de evasão para compor modelos de predição a fim de mitigar a evasão na EaD.

Além disso, como sugestão de trabalho futuro pode-se: verificar qual (quais) indicadores podem ter maior impacto na predição; identificar se há necessidade de utilizar todos os indicadores em modelos de predição; o porquê de um indicador utilizar dados que provêm de mais de um recurso do AVA; o porquê de um indicador utilizar dados de diferentes tipos (estruturado e não estruturado); investigar quais algoritmos são usados e como são construídos os modelos de predição relacionado os indicadores estudados; identificar se há um tratamento, considerando o dado usado com o algoritmo selecionado no modelo de predição, dentre outros.

Agradecimentos

Os autores gostariam de agradecer a CAPES - Código de Financiamento 001.

Referências

ABED. *Censo EAD.BR: relatório analítico da aprendizagem a distância no Brasil 2018 [livro eletrônico]/[organização] ABED – Associação Brasileira de Educação a Distância*. Curitiba: Intersaberes, 2019.

AGUDO-PEREGRINA Ángel F. *et al.* Can we predict success from log data in vles? classification of interactions for learning analytics and their relation with performance in vle-supported f2f and online learning. *Computers in Human Behavior*, v. 31, p. 542 – 550, 2014.

BALAKRISHNAN, G. *Predicting Student Retention in Massive Open Online Courses using Hidden Markov Models*, 2013. Available at: <http://www2.eecs.berkeley.edu/Pubs/TechRpts/2013/EECS-2013-109.html>.

BRITO, D. M. de *et al.* Identificação de estudantes do primeiro semestre com risco de evasão através de técnicas de data mining. In: *Congresso Internacional de Informática Educativa*, 2015. p. 459–463.

- CHEN, Y.; ZHANG, M. Mooc student dropout: Pattern and prevention. In: *ACM Turing 50th Celebration Conference*, 2017. p. 1–6.
- DIGIAMPIETRI, L. A.; NAKANO, F.; LAURETTO, M. d. S. Mineração de dados para identificação de alunos com alto risco de evasão: Um estudo de caso. *Revista de Graduação USP*, v. 1, n. 1, p. 17–23, 2016.
- HLIOUI, F.; ALOUI, N.; GARGOURI, F. Understanding learner engagement in a virtual learning environment. In: ABRAHAM, A. *et al.* (Ed.). *Intelligent Systems Design and Applications*. Cham: Springer International Publishing, 2020. p. 709–719.
- KHANNA, L.; SINGH, S. N.; ALAM, M. Educational data mining and its role in determining factors affecting students academic performance: A systematic review. In: *India International Conference on Information Processing*, 2016. p. 1–7.
- KIKA, C. A. *Supporting student experience management with learning analytics in the UK higher education sector*. Tese (Doutorado) — University of Bedfordshire, Inglaterra, 2018.
- KIZILCEC, R. F.; PÉREZ-SANAGUSTÍN, M.; MALDONADO, J. J. Self-regulated learning strategies predict learner behavior and goal attainment in massive open online courses. *Computers & Education*, v. 104, p. 18 – 33, 2017.
- LIMA, M. N.; ROBERTA, A. d. A. Educational datamining: A study of the factors that cause school dropout in higher education institutions in brazil. *RENOTE*, v. 18, n. 1, 2020.
- MACFADYEN, L. P.; DAWSON, S. Mining lms data to develop an “early warning system” for educators: A proof of concept. *Computers & education*, v. 54, n. 2, p. 588–599, 2010.
- MARQUES, L. T. *et al.* Mineração de dados auxiliando na descoberta das causas da evasão escolar: Um mapeamento sistemático da literatura. *RENOTE*, v. 17, n. 3, p. 194–203, 2019.
- MAZZA, R.; DIMITROVA, V. Coursevis: A graphical student monitoring tool for supporting instructors in web-based distance courses. *International Journal of Human-Computer Studies*, v. 65, n. 2, p. 125 – 139, 2007.
- MORAIS, A. M. de. *Abordagem avaliativa multidimensional para previsão da evasão do discente em cursos on-line*. Tese (Doutorado) — Universidade Federal de Campina Grande, Paraíba, Brasil, 2018.
- PETERSEN, K. *et al.* Systematic mapping studies in software engineering. In: *International Conference on Evaluation and Assessment in Software Engineering*, 2008. p. 68–77.
- QUEIROGA, E.; CECHINEL, C.; ARAÚJO, R. Um estudo do uso de contagem de interações semanais para predição precoce de evasão em educação a distância. *Workshop do Congresso Brasileiro de Informática na Educação*, p. 1074–1083, 2015.
- ROMERO, C.; VENTURA, S.; GARCÍA, E. Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, v. 51, n. 1, p. 368–384, 2008.
- TAMADA, M. M.; NETTO, J. F. de M.; LIMA, D. P. R. de. Predicting and reducing dropout in virtual learning using machine learning techniques: A systematic review. In: *IEEE Frontiers in Education Conference*, 2019. p. 1–9.