

USO DE DATA MINING E LEARNING ANALYTICS PARA ANÁLISE E PREVENÇÃO À EVASÃO ESCOLAR: UM MAPEAMENTO SISTEMÁTICO DA LITERATURA

Resumo – Com o surgimento e potencialidade da Educação a distância e o uso frequente das tecnologias nesta modalidade, observa-se claramente que o processo de ensino e de aprendizagem nos ambientes educacionais passou por uma transformação, com esse crescimento outras preocupações também despertaram a necessidade de estudos e pesquisas aplicadas de alguns temas. Este trabalho destaca um desses temas a evasão escolar. Apesar da evasão escolar ser bastante complexa, dados mostram que os maiores índices de evasão estão associados a modalidade que usam as ações remotas e em EaD, superando os números apresentados pelo ensino presencial. Nesta perspectiva este trabalho apresenta um Mapeamento Sistemático da Literatura sobre o Uso de Data Mining e Learning Analytics para análise e prevenção à evasão escolar, em que se buscou identificar, além de um conhecimento acerca do tema, ferramentas tecnológicas que trate da coleta, processamento e análise de dados ligados a evasão escolar. Nesta fase da pesquisa identificou-se e quantificou-se os trabalhos encontrados em dois portais com amplitude de diversas bases de dados científicas, com o objetivo de responder algumas questões de pesquisa. Por fim, o mapeamento identificou que os principais trabalhos da área se concentram em estudar fatores relacionados ao tema proposto.

Palavras-chave: mapeamento sistemático da literatura. evasão escolar. Data Mining. Learning Analytics.

Introdução

A educação técnica e profissionalizante tem suas particularidades e desafios, principalmente pelo foco na formação do cidadão para o tão exigente mercado de trabalho, com as suas constantes mudanças e crescentes exigências. Cada área de atuação profissional tem diferentes características para que se possam atender aos diversos perfis de estudantes, dando possibilidades de acesso ao ensino e as tecnologias de forma igualitária, segundo RODRIGUES, Armindo J. Apud Lidia (2009), não deve ser tratada como uma abordagem tradicional onde era sinônimo de uniformização, mas numa abordagem de atenção a diversidade e a igualdade com respeito pelas diferenças e pelas necessidades individuais, desenvolvendo as potencialidades de cada aluno através de percursos individualizados de aprendizagem, respeitando as características e o ritmo de cada um.

O surgimento das redes de comunicação, mais conhecida como Internet, iniciou uma nova etapa no rumo e tendência que o processo de aprendizagem tomou, confirmando assim o conceito de ciberespaço que, segundo Lévy (1999), é um ambiente de comunicação que surge da interligação de computadores em escala global e que envolve não apenas a infraestrutura material de comunicação digital, mas também as informações, e ainda envolvem às técnicas, às práticas, às atitudes, aos modos de pensamento e de valores que se desenvolvem juntamente com o ciberespaço.

Neste cenário, a educação a distância apresenta-se uma grande alternativa para a educação, pois as pessoas estão cada vez mais conectados e isso está mudando a forma como elas se relacionam com os outros, e a educação também é incluída nesse contexto, nota-se que as tecnologias atreladas a variedade de mídias digitais e as redes de comunicação estão

contribuindo cada dia mais com a educação na modalidade presencial quanto na educação a distância. As informações, cada vez mais acessíveis, tornam dispensáveis os processos de aprendizagem unidimensionais e com pouca interação, abrindo espaços para o uso de imagens, vídeos, sons, pesquisas e, principalmente, interatividade.

Segundo Silva (2013) a educação a distância é potencializada com o uso dos ambientes virtuais de aprendizagens (AVA), também conhecidos como Learning Management System (LMS), que são softwares e plataformas disponibilizadas na internet e que agregam ferramentas para criação, gestão e tutoria de atividades. Do mesmo modo, cita ainda que o MOODLE, sigla em inglês para Modular Object-Oriented Dynamic Learning Environment, ou seja, Ambiente de Aprendizado Modular Orientado ao Objeto, concebido por Martin Dougiamas em 1999 na Curtin University of Technology, o MOODLE vem sendo utilizado amplamente. Esse sucesso está relacionado por ser um sistema modular e customizável com possibilidades de instalação de novos recursos e por possuir características como flexibilidade de tempo e espaço, repositório de dados, espaços de interação e de aprendizagem (Moodle, 2019).

Conforme já citado a educação a distância é sem dúvida uma importante ferramenta de ensino, pois facilita o acesso de muitas pessoas à educação, transpassando barreiras físicas e temporais, contudo é notório o alto percentual de alunos que desistem ou evadem antes mesmo da percepção e intervenção dos gestores; isso é ratificado no texto de (ALVES, 2009), ele menciona que mesmo com todas as facilidades disponíveis: atrativa, motivadora e agradável que possa ser a aula no AVA, o processo apresenta consideráveis percentuais de evasão.

De acordo com o Ministério da Educação - MEC (BRASIL, 1996), a evasão escolar pode ser definida de três formas:

- a) evasão de curso: quando o estudante se desliga do curso em situações diversas tais como: abandono (deixa de matricular-se), desistência (oficial), transferência ou reopção (mudança de curso), exclusão por norma institucional;
- b) evasão da instituição: quando o estudante se desliga da instituição na qual está matriculado;
- c) evasão do sistema: quanto o estudante abandona de forma definitiva ou temporária o ensino.

Para Marun (2008), os alunos prestes a abandonar os estudos normalmente o farão sem qualquer tipo de pronunciamento dos responsáveis pelo processo educativo, se, no caso, forem maiores de idade. Não é menos comum, no entanto, vê-los retornar à instituição da mesma forma que a deixaram, sem justificativas plausíveis para a sua ausência temporária.

A evasão é apontada por grande parte das instituições como um dos maiores obstáculos enfrentados nos diferentes tipos de cursos EAD. Em todos os tipos de curso, nenhuma instituição apontou taxas de evasão superiores a 75% e, na maioria dos casos, a evasão identificada se concentra na faixa de até 25%. No caso dos cursos regulamentados semipresenciais, 36

instituições afirmaram que a evasão foi de até 25%; 24, que foi de 26% a 50%, e apenas 5 declararam que a evasão foi superior a 50%. Censo EAD.BR (2014).

Ciente disso, este trabalho de pesquisa contribuirá para que possamos atender as diversas realidades que chegam em nossa instituição educacional, melhorando nossos conhecimentos e técnicas através do desenvolvimento de uma ferramenta de auxílio para tomada de decisões sobre os pilares da evasão. Para isso faz-se necessário que a ferramenta citada contribua para uma gestão proativa e preventiva atuando diretamente na coleta, processamento e apresentação das informações de acesso dos discentes, extraídos do ambiente virtual de aprendizagem na plataforma MOODLE.

Objetivo

O objetivo deste mapeamento é apresentar uma visão geral de pesquisas que vem sendo conduzidas sobre o tema evasão escolar e as diversas formas de análise de dados através de mineração e Learning Analytics, aplicada a educação.

Metodologia

Segundo (Petersen et al. 2008) os mapeamentos sistemáticos são um método que se propõe em organizar por categoria uma grande quantidade de estudos existentes na literatura com base em seus resultados, contabilizando as contribuições a partir desta categorização. Esses mapeamentos buscam identificar lacunas e oportunidades de pesquisa, com o objetivo aumentar os conhecimentos do pesquisador sobre o tema abordado (Budgen et al. 2008).

Diante desta realidade, a metodologia aplicada para o mapeamento desenvolvido foi baseada no processo descrito por (Petersen et al. 2008), segundo o autor, há cinco passos essenciais a serem seguidos:

1. Definir as questões de pesquisa;
2. Criar as strings de busca;
3. Escolher as bases de dados utilizadas na pesquisa;
4. Analisar os trabalhos selecionados;
5. Extrair os dados e concluir o mapeamento.

Análise de dados

Avaliando que as questões de pesquisa devem elucidar os objetivos do estudo de mapeamento, as seguintes questões foram elaboradas:

QP₁: Quais os principais motivos que influenciam no índice de evasão?

QP₂: Quais são as técnicas utilizadas em Learning Analytics atrelado a evasão?

QP₃: Quais as ferramentas de Mineração de Dados utilizadas para análise da evasão escolar?

QP₄: Quais as técnicas de Mineração de Dados têm sido utilizadas para a descoberta das causas da evasão escolar?

Inicialmente foram realizadas algumas buscas com as palavras que seriam utilizadas como palavras-chave no processo da pesquisa. Com o uso de combinações das palavras *Learning Analytics*, *evasion*, *school* e *dropout*, todas no idioma inglês. Conforme mostrado na Strings de Busca abaixo:

("Data Mining" OR "data extraction" OR "Learning Analytics") AND ("school dropout" OR "school drop-out" OR "school evasion" OR "school truancy" OR "school failure")

Para este Mapeamento Sistemático da Literatura, as bases selecionadas foram as seguintes: Scopus, IEEEExplore, Oxford Journals, Science Citation Index Expanded e ERIC (U.S. Dept. of Education). O critério de escolha dessas bases de dados se deve ao fato de serem as bases mais difundidas no campo da tecnologia. Essas bases de dados foram desenvolvidas com o intuito de propiciar, em uma única plataforma, centenas de revistas científicas e seus respectivos trabalhos, assentes por qualidade e originalidade (Dantas et al., 2018).

Conforme afirma (Fuzeto e Braga, 2016) os critérios de inclusão e exclusão são definidos para auxiliar na condução de um Mapeamento Sistemático da Literatura, com o intuito de apoiar a classificação de relevância dos estudos. Tais critérios estão diretamente relacionados às questões de pesquisa em análise. Para o mapeamento foram definidos os seguintes critérios de inclusão (CI) e exclusão (CE):

CI₁: Trate do tema Mineração de Dados e Learning Analytics ligados a evasão;

CI₂: Estivessem escritos em inglês ou português;

CI₃: Das bases: Scopus, IEEEExplore, Oxford Journals, Science Citation Index Expanded e ERIC (U.S. Dept. of Education);

CE₁: Trabalhos em andamento, ou seja, que tenham sido concluídos;

CE₂: Estudos anteriores a 2000;

CE₃: Artigos em outros idiomas que não o inglês ou português;

CE₄: Serão excluídas algumas bases após a consulta da Strings.

Foram seguidas 3 (três) etapas de filtragem de trabalhos, conforme descreve-se a seguir:

1ª Etapa: Aplicação das *strings* de busca;

2ª Etapa: Aplicação dos critérios de inclusão e exclusão;

3ª Etapa: Estudos mais relevantes para a pesquisa dos temas proposto.

Na condução do Mapeamento Sistemático da Literatura o trecho a seguir detalha o procedimento de busca e seleção de trabalhos nas bases pautadas nos critérios de inclusão. As buscas nas bases de dados foram realizadas por meio de buscas avançadas, na base da CAPES e IEEEXplor, sofrendo a *string* de busca pequenos ajustes.

Bases	Etapa 01	Etapa 02	Etapa 3
Scopus	154	12	5
IEEEXplore	9	3	1
Oxford Journals	8	2	1
Science Citation Index Expanded	73	7	2
ERIC (U.S. Dept. of Education)	57	4	1
Total	301	28	10

Tabela 1. – Resultados das buscas. (Fonte: Autoria própria)

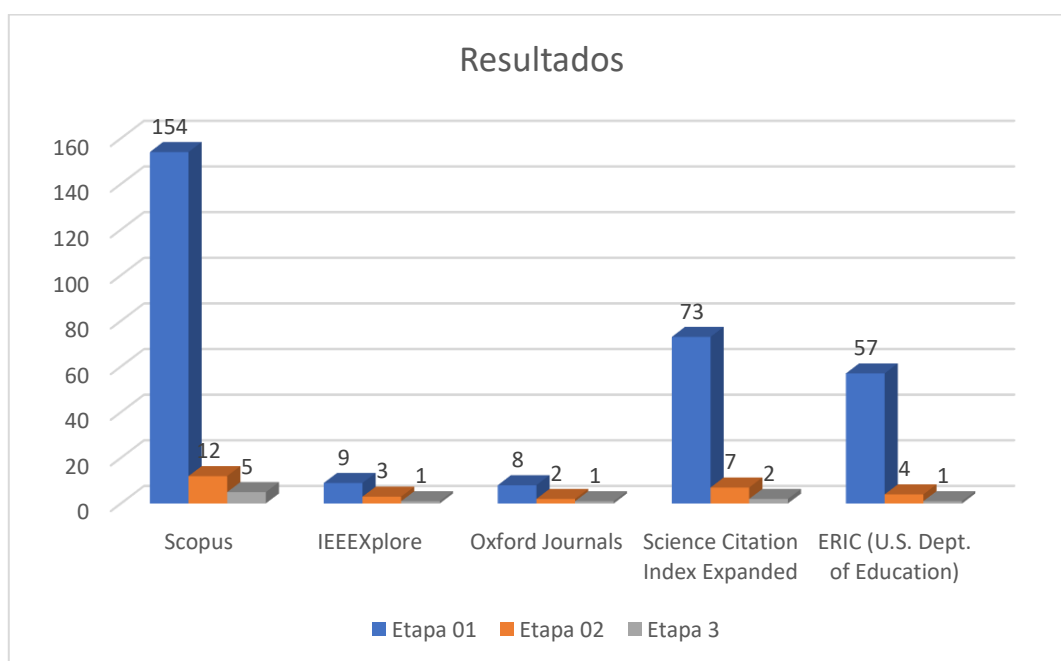


Figura 1. Representação gráfica dos resultados das buscas. (Fonte: Autoria própria)

Observa-se no gráfico, figura 1, e nos dados tabulados, tabela 1, os trabalhos analisados em cada etapa. Na primeira etapa foram encontrados 301 trabalho, ressalto que muitos trabalhos são repeditos, após aplicação dos critérios de inclusão (CI) e dos critérios de exclusão (CE), o número apresentado foi 28, quantitativo que já facilitou a análise dos estudos mais relevantes

para a pesquisa, onde foram selecionados dez (10) trabalhos. Após seguir os passos da pesquisa que pode-se com o resultado responder as questões de pesquisa, conforme descrito abaixo:

QP₁: Quais os principais motivos que influenciam no índice de evasão?

Como base nos 10 documentos lidos, foi possível delimitar as questões macro dos motivos da evasão, destacam-se: fatores internos, fatores externos e características individuais dos estudantes.

QP₂: Quais são as técnicas utilizadas em Learning Analytics atrelado a evasão?

Nas literaturas pesquisadas as técnicas de análise de dados baseadas em Mineração de Dados e Aprendizagem de Máquina destacam-se: os modelos baseados em Árvores de Decisão (Decision Trees - DT) [Almeda et al. 2018; Pardo et al. 2016; Jayaprakash et al. 2014; Romero et al. 2013], Aprendizagem Profunda (Deep Learning - DL) [Okubo et al. 2017], Máquinas de Vetor de Suporte (Support Vector Machines - SVM) [Jayaprakash et al. 2014] e Métodos Bayesianos (Bayesian Methods - NB).

QP₃: Quais as ferramentas de Mineração de Dados utilizadas para análise da evasão escolar?

Com base no estudo foi possível verificar que as ferramentas Weka e Mplus destacam-se, pois foram dissertadas em em três (3) trabalhos. A Weka foi utilizada nos trabalhos de Márquez-Vera et al. (2016), Márquez-Vera et al. (2013) e Jiménez-Gómez et al. (2015). O Mplus foi utilizado nos trabalhos de Parr e Bonitz (2015), Archambault et al. (2009) e Fall e Roberts (2012).

QP₄: Quais as técnicas de Mineração de Dados têm sido utilizadas para a descoberta das causas da evasão escolar?

As técnicas de Mineração de Dados que foram exploradas nos trabalhos encontrados. Destacam-se as técnicas de Classificação que foram utilizadas em quatro (3) trabalhos, em seguida as Equações Estruturais empregadas em três (2) trabalhos.

A técnica de Classificação foi dissertada nas publicações de Rovira et al. (2017), Márquez-Vera et al. (2016) e Jiménez-Gómez et al. (2015). As Equações Estruturais utilizadas nos trabalhos de Parr e Bonitz (2015) e Archambault et al. (2009).

Conclusão

Neste trabalho foi apresentado os resultados de um Mapeamento Sistemático da Literatura sobre a evasão escolar, em que se buscou levantar as técnicas de Learning Analytics e mineração de dados, ferramentas de mineração de dados e os principais motivos atrelado ao tema.

A partir dos resultados, evidenciou-se que, através de aplicação de técnicas e algoritmo a ferramenta Weka se consolida a mais utilizada e que a técnica de classificação, vem sendo amplamente utilizada com alto índice de precisão de análise na tendência de evasão.

Por fim, é importante destacar que qualquer pesquisa desta natureza possui riscos à validade dos seus resultados. Trazendo em si pontos de fragilidade, tais como, o pouco tempo para pesquisa e leitura dos trabalhos encontrados, o que pode ter causado lapsos de apresentar trabalhos mais relevantes, ficando a necessidade de um estudo mais completo para consolidar este trabalho.

Referências

- ALVES, L.; **Um olhar pedagógico das interfaces do Moodle**. In: Lynn; Barros, Daniela; Okada, Alexandra (Org.). MOODLE: Estratégias Pedagógicas e Estudos de Caso. Salvador: Eduneb, p 187-201. 2009
- BRASIL. Ministério da Educação. **Diplomação, retenção e evasão nos cursos de graduação em instituições de ensino superior públicas**. Brasília,1996.
- Budgen, D., Kitchenham, B. A., Charters, S., Turner, M., Brereton, P. & Linkman, S. (2008), **'Presenting software engineering results using structured abstracts: A randomised experiment'**, *Empirical Software Engineering* . Accepted for publication.
- Gamma, E., Helm, R., Johnson, R. & Vlissides, J. (1995), **Design Patterns: Elements of Reusable Object-Oriented Software**,Addison-Wesley.
- Budgen, D., Kitchenham, B. A., Charters, S., Turner, M., Brereton, P. & Linkman, S. (2008), **'Presenting software engineering results using structured abstracts: A randomised experiment'**, *Empirical Software Engineering* . Accepted for publication. Gamma, E., Helm, R., Johnson, R. & Vlissides, J. (1995), *Design Patterns: Elements of Reusable Object-Oriented Software*,Addison-Wesley.
- CENSO EAD. BR. **Relatório Analítico da Aprendizagem a distância no Brasil**. 1ª edição, 2015.,p.157.2014.
- Dantas, Á. M. C., Viana, H., Abijaude, J., and Sobreira, P. (2018). **Internet das coisas e aprendizagem colaborativa: Revisão sistemática da literatura**. In Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE), volume 29, page 278.
- DINIZ, Ubiratan Morais. **EVASÃO DO CURSO DE GRADUAÇÃO A DISTÂNCIA:Um estudo do Bacharelado em Administração Publica na UFPA**. Belém-Pará, 2013. (Dissertação de mestrado) Disponível em: <http://www.repositorio.ufpa.br/jspui/handle/2011/4625>. Acesso em 02/11/2020.
- FRANCO, C.P.. **Possibilidades de participação na sala de aula virtual: uma análise dos modos de interação em um componente on-line de leitura instrumental em inglês**. *Eutomia - Revista Online de Literatura e Linguística*, 3(1), 1-15, 2010. Disponível em <http://www.claudiofranco.com.br/textos/franco_eutomia.pdf>. Acesso em: 01/11/2020.

FUZETO, R. and BRAGA, R. (2016). **Um mapeamento sistemático em progresso sobre internet das coisas e educação à distância.** In Anais dos Workshops do Congresso Brasileiro de Informática na Educação, volume 5, page 1334.

GAIOSO, N. d. L. (2005). **O fenômeno da evasão escolar na educação superior no brasil.** Brasília, DF: Universidade Católica de Brasília.

LÉVY, Pierre. **Cibercultura.** São Paulo: Atlas, 1999.

LIMA, E.S. **Revisão Sistemática da Literatura: as causas da evasão nos cursos de graduação a distância.** Revista Educaonline, v.9, no 3, setembro a dezembro, p.32-69. 2015.

MARUN, Dulcinéia Janúncio. **Evasão Escolar no Ensino Médio: Um Estudo Sobre Trajetórias Acidentadas.** Pontifícia Universidade de São Paulo, 2008.

MOODLE. 2019. Disponível em: < <https://moodle.org> >. Acesso em: 01 de novembro de 2021.

PETERSEN, K., Feldt, R., Mujtaba, S., and MATTSSON, M. (2008). **Systematic mapping studies in software engineering.** In Ease, volume 8, pages 68–77.

SILVA, Lidia Martins. Monografia, **EDUCAÇÃO INCLUSIVA E FORMAÇÃO DE PROFESSORES, 2009.** Disponível em: <http://bento.ifrs.edu.br/site/midias/arquivos/2010069353641lidia_monografia.pdf> Acesso em: 03/11/2020.

SILVA, Robson S. da. **Moodle para autores e tutores.** 3ª ed. São Paulo: Novatec Editora, 2013.

FILHO, Ruy Leite Berger. **Educação profissional no Brasil: novos rumos, 1999.** Disponível em: <<https://rieoei.org/historico/documentos/rie20a03.htm>> Acesso em: 04 de novembro de 2021.

SOUSA, A. de A. **Evasão escolar no ensino médio: velhos ou novos dilemas?**, 2011. Disponível em: < <http://essentiaeditora.iff.edu.br/index.php/vertices/article/viewFile/1220/641> > . Acesso em 04 de novembro de 2021.