

Mineração de Dados Educacionais para Aperfeiçoamento da Gestão do *E-learning*: uma Revisão Sistemática de Literatura

Educational Data Mining to Improve E-learning Management: a Systematic Literature Review

Humberto Rodrigues Marques

Universidade Federal de Lavras – UFLA - Brasil
hbetorm@hotmail.com
ORCID: 0000-0003-0557-8242

Larissa Garcia Gomes

Universidade Federal de Lavras – UFLA - Brasil
larissa.gomes@posgrad.ufla.br
ORCID: 0000-0003-1375-1383

André Luiz Zambalde

Universidade Federal de Lavras – UFLA - Brasil
zambaufla@gmail.com
ORCID: 0000-0003-0129-7451

André Grützmann

Universidade Federal de Lavras – UFLA - Brasil
andregrutzmann@gmail.com
ORCID: 0000-0003-3568-5655

Submetido em 11/09/2019; **Aprovado** em 13/07/2020.

Resumo

Objetivo: Uma vez que a mineração de dados educacionais auxilia na análise da grande quantidade de dados que sistemas de *e-learning* geram, aperfeiçoando os sistemas de ensino-aprendizagem, o presente estudo procura, por meio de uma revisão sistemática de literatura, verificar como o tema está sendo abordado nas pesquisas do campo nos últimos anos. **Metodologia:** Dado que tal contexto é emergente, novo e dinâmico e que necessita de análises de sua estrutura para entender suas características, dificuldades e potencialidades, este estudo foi desenvolvido para analisar as estruturas existentes de sistemas de *e-learning* e identificar como o processo de mineração de dados educacionais pode ser aplicado no ensino a distância, por meio de uma revisão sistemática de literatura. **Resultados:** Os resultados estão divididos em duas partes: (i) análise descritiva, em que os artigos são apresentados a partir de uma análise bibliométrica, e (ii) análise semântica, em que os artigos são analisados a partir do modelo de quatro etapas do processo de mineração de dados educacionais. **Contribuições:** Os resultados contribuem com recomendações práticas e acadêmicas, pois apresentam tanto como o processo de mineração de dados interfere na gestão do *e-learning* quanto um panorama das pesquisas existentes, a fim de conhecer a pesquisa atual da temática analisada.

Palavras-chave: Ensino a Distância. Educação. Administração Educacional. Instituições Educacionais.

Abstract

Purpose: Since the mining of educational data helps to analyze the large amount of data generated by e-learning systems, improving the teaching-learning systems, the present study seeks, through a systematic literature review, to check how the topic has been, in recent years, researched. **Approach:** Once this context is new, emerging and dynamic and that needs analysis of its structure to understand its dynamics, difficulties and potential, this study was developed to analyze the existing structures of e-learning systems, as well as to identify how the educational data mining process can be applied in e-learning. **Results:** The results are divided into two parts: (i) descriptive analysis, in which the articles are presented from a bibliometric analysis, and (ii) semantic analysis, in which the articles are analyzed

using the four-step model of the educational data mining process. **Contributions:** The results contribute with practical and academic recommendations, as they present as much as the process of data mining interferes in the management of e-learning, as well as an overview of the existing researches, in order to know the current research of the analyzed theme.

Keywords: E-learning. Education. Educational administration. Educational Institutions.

1. Introdução

Nos últimos anos, o *e-learning* tornou-se um dos mais populares modos de entrega de conhecimento, principalmente devido aos avanços e à maior acessibilidade à tecnologia (Lamph et al., 2018), de forma que se tornou uma tendência de aprendizagem que está crescendo rapidamente e continua a influenciar o ensino e a aprendizagem na sala de aula e, até mesmo, além dela (Male & Pattinson, 2011). Assim, o *e-learning* pode ser considerado como o uso de tecnologia de telecomunicações para fornecer informações para educação e treinamento, emergindo como o paradigma da educação moderna (Khadabandeh, Afshari & Manian, 2010).

Frequentemente, quando se trata de *e-learning*, apesar de considerar-se que é suficiente “integrar” a solução tecnológica no processo de ensino e aprendizagem, pode-se verificar que a dinâmica é muito mais complexa e deve ser adequadamente (re)projetada (Ćukušić et al., 2010). Segundo Oliveira, Cunha e Nakayama (2016), é essencial estabelecer estratégias e mecanismos pelos quais se possa garantir que o sistema funcionará efetivamente como pretendido, dado que o *e-learning* é composto de vários componentes que devem operar de maneira integrada.

Apesar de o *e-learning* ainda estar em sua infância, observa-se que está se desenvolvendo rapidamente, fato que levantou diferentes questões acerca de sua qualidade (Farid et al., 2018), de modo que a relação entre aprendizagem e inovação em tecnologias é agora uma área importante a ser estudada (Haq et al., 2018). Para Male e Pattinson (2011), alguns dos desafios do desenvolvimento de um sistema de *e-learning* são a identificação de componentes apropriados e relevantes do sistema, de suas inter-relações e de seu papel como parte do sistema.

Dado que o aumento de recursos de *e-learning* e de bancos de dados de alunos leva a enormes repositórios (Tan et al., 2017), recentemente, o impacto do ensino virtual começou a ser analisado usando técnicas de “mineração de dados” (MD), um sistema de informação baseado em computador e dedicado a escanear enormes fontes de dados, gerar informações e descobrir um conhecimento mais profundo, além de estabelecer novos *links* entre grandes quantidades de dados (Chamizo-Gonzalez et al., 2015). Já a mineração de dados educacionais (MDE), que se refere à aplicação da MD ao campo educacional, está surgindo como um novo campo de pesquisa (Lee, 2017). De acordo com a sociedade internacional de MDE, é uma disciplina emergente, preocupada em desenvolver métodos para explorar dados que vêm de ambientes educacionais e em usar tais métodos para entender melhor os alunos e o modo como aprendem (Joorabchi, English & Mahdi, 2016).

Segundo Peña-Ayala (2014), verificam-se tendências relacionadas à MDE, tais como a integração à arquitetura de sistemas educacionais baseados em computadores e a análise de suas várias características. Diante do exposto, apurada a necessidade de utilizar a MDE em contextos de *e-learning* como forma de corroborar no aperfeiçoamento do novo método de ensino-aprendizagem, este estudo objetiva, por meio de uma revisão sistemática de literatura, verificar como tais temas estão sendo abordados nas pesquisas do campo nos últimos anos.

A revisão sistemática foi desenvolvida visando a analisar as estruturas existentes de sistemas de *e-learning* e identificar como o processo de MDE está sendo aplicado no ensino a distância. Dado que o contexto é emergente, novo e dinâmico, analisar sua estrutura é fundamental para entender suas características, dificuldades e potencialidades. Tais informações podem ser úteis para especialistas do setor, pois será apresentado como o processo de MDE interfere na gestão do *e-learning*, e para estudiosos da área, pois será apresentado um panorama das pesquisas existentes, a fim de conhecer a pesquisa atual na temática analisada.

O trabalho está estruturado em mais seis partes, além da introdução. Na segunda seção serão apresentados os fundamentos teóricos. Em seguida serão destacados alguns trabalhos relacionados publicados na literatura. Na quarta seção serão apresentados os procedimentos metodológicos. Posteriormente, serão apresentados os resultados, divididos em duas partes: (i) análise descritiva e (ii) análise semântica. Por fim, as duas últimas seções são, respectivamente, a conclusão e as referências.

2. Mineração de dados educacionais no *e-learning*

Nos últimos anos, com a proliferação de tecnologias de comunicação para a aprendizagem nas escolas e nos sistemas de ensino superior, a aprendizagem colaborativa on-line tornou-se comum nos processos de ensino e aprendizagem (Weiser, Blau & Eshet-Alkalai, 2018), de modo que as instituições acadêmicas precisaram reformular e transformar a forma de ensinar, trabalhar e conduzir pesquisas (Harrati, Bouchrika & Mahfouf, 2017).

Dado o contexto, o ensino a distância, estimulado pela internet, tornou-se um dos principais componentes do ensino superior, eliminando a barreira da distância, uma vez que entrega os materiais do curso sem a necessidade de aluno e professor estarem em um mesmo ambiente físico (Tan et al., 2017). Com a proposta do *e-learning*, verifica-se que a dinâmica de aprendizagem mudou, de modo que instrutores e professores não podem mais observar os alunos através das salas de aula, assim como os alunos também não podem mais ter comunicação cara a cara com instrutores e outros alunos, dado que as atividades são realizadas por plataforma on-line (Luo, Zhang & Qi, 2017).

Usado inicialmente como um termo genérico para fornecer educação flexível, com origem no Reino Unido, o *e-learning* passou por estágios evolutivos, de forma que hoje requer o gerenciamento de sistemas complexos com o auxílio de tecnologias de mídia (Simo, Barbulescu & Kilyeni, 2015). O ensino a distância é, hoje, essencial para a aprendizagem em uma sociedade moderna, sendo utilizado tanto por empresas no treinamento e desenvolvimento de seus funcionários como por universidades na formação de estudantes (Khodabandeh et al., 2010).

Embora o *e-learning* tenha atraído muita atenção nos últimos anos e haja uma grande quantidade de informações disponíveis sobre ele, o desenvolvimento de material de conteúdo de alta qualidade para o ensino a distância ainda é caro e consome tempo (Siqueira, Braz & Melo, 2007). A falta de uma boa estrutura apresenta sérios desafios metodológicos para os praticantes de *e-learning* (Haq et al., 2018), de modo que cabe aos agentes de *software* lidar com as limitações dos sistemas eletrônicos, apoiando processos de aprendizagem que visam e fornecem materiais *just-in-time* para os alunos, dando suporte a instrutores, especialistas e alunos com o *design* e a entrega do curso, e personalizando os materiais de acordo com as características dos alunos (Gregg, 2007).

Dado que os cursos on-line ainda são baseados em materiais de aprendizagem estáticos, que não consideram a diversidade dos alunos, o desenvolvimento de sistemas educacionais integrados tem sido visto como uma solução para ambientes de aprendizado individual mais ricos, oferecendo educação personalizada aos alunos (Romero & Ventura, 2007). Assim sendo, uma maneira de melhorar a qualidade dos processos educacionais é aperfeiçoando a tomada de decisão, fornecendo à administração conhecimentos úteis que podem ser descobertos a partir de bancos de dados educacionais ou em formulários de alunos e professores, que podem ser extraídos através da tecnologia de mineração de dados (Chalaris et al., 2014).

Em sua essência, a mineração de dados é uma abordagem indutiva, analítica e exploratória (Angeli et al., 2017), que busca descobrir padrões de dados, organizar informações de relacionamentos ocultos, estruturar regras de associação, estimar valores de itens desconhecidos para classificar objetos, compor *clusters* de objetos homogêneos e desvendar muitos tipos de descobertas que não são facilmente conhecidos, caracterizando um suporte valioso para a tomada de decisão (Peña-Ayala, 2014).

A mineração de dados tem sido usada há muito tempo em marketing, publicidade, saúde, engenharia e sistemas de informação (Angeli et al., 2017), além de muitos outros setores profissionais, como bancos, seguradoras e empresas comerciais, para aprofundar as informações de clientes e adequar seus produtos ou serviços aos seus requisitos (Chamizo-Gonzalez et al., 2015). Quanto ao contexto educacional, uma forma específica da mineração de dados é conhecida como mineração de dados educacional (MDE) (Tan et al., 2017), utilizada para a descoberta de conhecimento, tomada de decisão e estabelecimento de ligações entre alunos, aluno-professor e ensino-aprendizagem (Chamizo-Gonzalez et al., 2015; Asif, Merceron, Ali & Haider, 2017).

Nos últimos 10 anos, a MDE emergiu como uma área distinta de pesquisa, preocupada com o uso de técnicas para responder questões educacionais (Angeli et al., 2017). Por exemplo, técnicas de MDE são eficazes para identificar precocemente falhas acadêmicas dos alunos e fornecer aos educadores ou professores informações relevantes para a tomada de decisão (Costa et al., 2017), uma vez que o MDE procura entender os estudantes e seus estilos de aprendizagem de forma mais abrangente, em um esforço para elaborar políticas educacionais que melhorem seu desempenho acadêmico e reduzam as taxas de reprovação no final de cada ano letivo (Fernandes et al., 2019).

O estudo de Romero, Ventura e García (2008) descreve o passo a passo do processo de MDE, bem como as principais técnicas de mineração de dados existentes, enfatizando que a aplicação da mineração de dados em sistemas de *e-learning* é um ciclo iterativo, em que o conhecimento extraído deve entrar no circuito do sistema e orientar, facilitar e melhorar o aprendizado como um todo, não apenas transformando dados em conhecimento, mas também filtrando o conhecimento extraído para a tomada de decisão. Para tanto, o processo de mineração de dados de *e-learning* consiste em quatro etapas, sendo elas, segundo Romero et al. (2009):

- 1) Coletar dados: o sistema de gerenciamento de curso (SGC) é usado pelos alunos, e as informações de uso e interação são armazenadas em banco de dados e depois coletadas.
- 2) Pré-processar os dados: os dados são limpos e transformados em um formato apropriado para serem extraídos.
- 3) Aplicar mineração de dados: os algoritmos de mineração de dados são aplicados para construir e executar o modelo que descobre e resume o conhecimento de interesse para o usuário (instrutor, aluno e administrador).
- 4) Interpretar, avaliar e implantar os resultados: os resultados ou modelos obtidos são interpretados e usados pelo instrutor para outras ações.

3. Trabalhos relacionados

Especificamente sobre o tema de MDE, alguns trabalhos foram desenvolvidos na área. A seguir, apresentam-se alguns destes estudos com uma breve contextualização.

Romero e Ventura (2010), ao identificarem que o campo da MDE estava crescendo, desenvolveram uma revisão de literatura para entenderem o campo, ao mesmo tempo em que propuseram linhas de pesquisas para o futuro. Os autores enfatizaram que a MDE, na época, estava se aproximando da adolescência, ou seja, não estava mais em seus primeiros dias, mas ainda não era uma área madura. Também de uma forma mais ampla, Papamitsiou e Economides (2014) analisaram 40 estudos publicados entre 2008 e 2013, identificando as questões de pesquisa, a metodologia e os principais resultados dos artigos encontrados, destacando, principalmente: (i) quais são os objetivos básicos de pesquisa dos artigos (em termos de métricas mensuráveis) e quais métodos os pesquisadores seguiram para alcançar os objetivos; e (ii) quais as tecnologias de pesquisa emergentes exploradas nos artigos.

O trabalho de Mohamad e Tasir (2013) teve como objetivo examinar como a mineração de dados foi abordada por estudiosos anteriores e as últimas tendências sobre mineração de dados na pesquisa educacional, tendo como principal resultado a constatação de que a maioria das pesquisas sobre MDE prestam muita atenção ao uso de *e-learning* em sistemas como Moodle, WebCT, Blackboard e que, num futuro próximo, deveriam mudar o foco do *e-learning* para o uso de ferramentas de redes sociais como Blog e Facebook, já que esses aplicativos possuem grande popularidade entre os alunos e são adequados para o aprendizado colaborativo.

Mais especificamente, Dutt, Ismail e Herawan (2017) analisaram, por meio de uma revisão sistemática da literatura, o algoritmo de *clustering* e sua aplicabilidade e usabilidade no contexto do MDE, dado que muitos algoritmos tradicionais de mineração de dados não podem ser aplicados diretamente a problemas educacionais. Já no trabalho de Slater et al. (2017), os autores destacaram algumas das ferramentas mais amplamente utilizadas, acessíveis e poderosas disponíveis para o pesquisador interessado em realizar pesquisas sobre MDE, além de informações mais descritivas, como preço e facilidade de uso.

O estudo de Shahiri e Husain (2015) teve como objetivo fornecer uma visão geral sobre as técnicas de mineração de dados que foram usadas para prever o desempenho dos alunos, dado que é possível melhorar a performance e o sucesso discente de maneira mais eficaz usando técnicas de MDE, trazendo benefícios e impactos para estudantes, educadores e instituições acadêmicas. Já Rodrigues, Isotani e Zárate (2018) identificaram, por meio de uma revisão, os principais meios de avaliação em sistemas de *e-learning*, uma vez que, segundo eles, são necessários novos mecanismos para avaliar a eficácia pedagógica destes sistemas de ensino.

O objetivo do trabalho de Manjarres, Sandoval e Suárez (2018) foi apresentar uma revisão dos trabalhos nos quais as técnicas de mineração de dados foram utilizadas para solucionar problemas específicos da educação e fazer uma classificação associada a diversos cenários nos quais foram aplicadas. Os autores identificaram mais de doze técnicas de MDE que podem ser usadas para o *e-learning*, principalmente em problemas como a identificação de padrões de aprendizagem, de padrões dos alunos e,

mais recentemente, do abandono escolar. A questão do abandono foi tema do trabalho de Agrusti, Bonnavolontà e Mezzini (2019), que utilizaram procedimentos de revisão para identificar estudos que utilizam técnicas educacionais de mineração de dados (técnicas, algoritmos e ferramentas) para prever o abandono de alunos em cursos tradicionais de universidades.

No trabalho de Liñán & Pérez (2015), os autores estudaram as semelhanças e as diferenças, além da origem, dos objetivos, da evolução e dos desafios, entre MDE e análise de aprendizagem, dois campos de pesquisa relativamente novos e populares, relacionando estes temas à coleta, à análise e à interpretação de dados sobre educação. No mesmo contexto, Aldowah, Al-Samarraie e Fauzy (2019) realizaram uma revisão abrangente da MDE e da análise de aprendizagem no ensino superior, de modo que identificaram técnicas de mineração de dados mais e menos utilizadas para adaptar o processo de aprendizagem às necessidades individuais dos alunos.

Diante das revisões apresentadas, verificamos que a maioria se concentra em alguns assuntos específicos e dá mais ênfase à análise de algumas etapas do processo de MDE. Diferentemente, como apresentado no presente estudo, procura-se abordar aqui a temática da MDE a partir das quatro etapas propostas por Romero et al. (2008), como forma de proporcionar um panorama não apenas do campo de estudo, como também de todas as etapas do processo de MDE. Ainda, este estudo foca na MDE para cursos de *e-learning*, ou seja, a revisão abarca a aplicação da MDE em cursos totalmente on-line e não apenas em banco de dados com informações de alunos.

4. Procedimentos Metodológicos

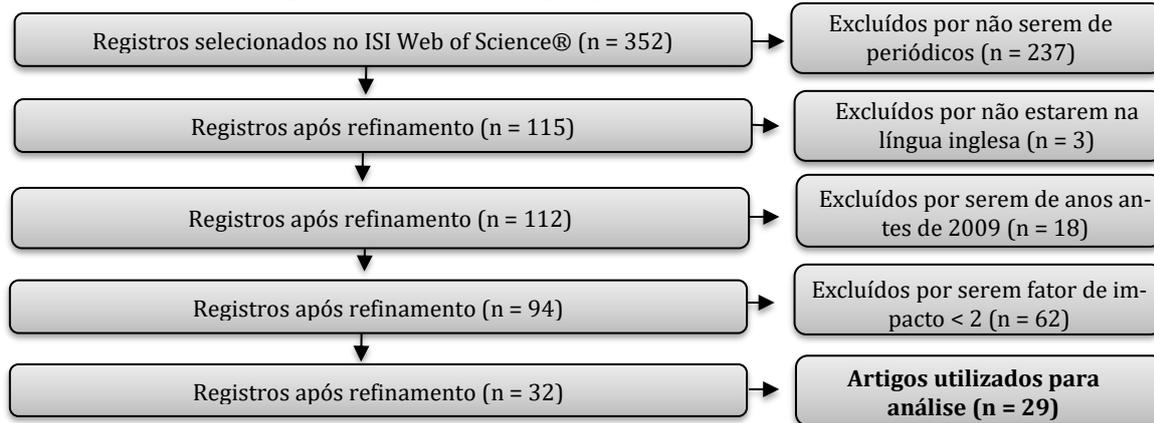
Como forma de atingir o objetivo proposto, realizou-se uma revisão sistemática de literatura, metodologia que procura analisar a pluralidade de conhecimento existente numa determinada área acadêmica, sendo considerada transparente e passível de ser replicada (Tranfield, Denyer & Smart, 2003; Galvão, Sawada & Trevizan, 2004). Como forma de expor a viabilidade da pesquisa, a presente seção apresenta seu delineamento metodológico por meio de quatro fases: (1) procedimentos de busca e seleção; (2) critérios de seleção; (3) análise descritiva dos artigos; (4) análise semântica.

Os procedimentos de busca e seleção, que ocorreram no mês de junho de 2018, foram realizados no banco de dados ISI Web of Science® da Thomson Reuters, dado que o repositório possui uma grande abrangência de documentos, além de ser utilizado tanto pelo meio acadêmico quanto organizacional e possibilitar meios para exportar e trabalhar os dados. Definiu-se uma estratégia de busca que continha as palavras-chave “*Data Mining*” e “*Educational Data Mining*” combinadas com “*e-learning*”, empregada no campo de busca “tópicos” da base, que contempla aspectos dos artigos como Título, Resumo, Palavras-chave de autor e Palavras-chave da própria base de dados. A busca resultou em 352 artigos recuperados.

Com relação aos critérios de seleção (Figura 1), foram adotados os seguintes procedimentos de exclusão dos 352 documentos recuperados: (i) foram eliminados 237 artigos por não serem textos publicados em periódicos, como, por exemplo, artigos publicados em eventos, congressos, simpósios e afins; (ii) foram excluídos 3 artigos por não estarem publicados na língua inglesa; (iii) foram excluídos 18 artigos publicados antes de 2009, uma vez que o período de tempo adotado foram os últimos 10 anos, de 2009 a 2018, dado que apenas nos últimos anos os avanços tecnológicos deram assistência às plataformas e aos atores educacionais (Rodrigues et al., 2018), e apenas nos últimos 10 anos é que grandes avanços foram feitos no campo de pesquisa sobre MDE (Angeli et al., 2017; Bakhshinategh et al., 2018); (iv) foram excluídos 62 documentos, publicados em revistas com fator de impacto menor que 2; e, por fim, (v) após a leitura completa dos artigos finais, 3 trabalhos foram excluídos por não corresponderem ao escopo da pesquisa, dado que um não abordava MDE no seu desenvolvimento, outro empregava a perspectiva do *e-learning* na manufatura, e o último realizou uma revisão sistemática e não trazia resultados práticos que contribuíssem para a pesquisa.

A terceira fase consistiu na análise descritiva dos artigos selecionados na amostra final, como forma de caracterizar os documentos analisados. Os 29 artigos recuperados para análise foram tabulados em uma planilha eletrônica e sistematizados nas principais características dos artigos: (i) evolução de publicações; (ii) autores que mais publicaram no período; (iii) origem e área de pesquisa; (iv) revistas que mais publicaram; e (v) artigos mais citados da área.

Figura 1 - Desenho da pesquisa e critérios de seleção



Fonte: Elaborado pelos autores.

A quarta fase representa a análise semântica, em que foi realizada uma análise qualitativa dos artigos recuperados. Os 29 artigos selecionados foram analisados com base no estudo de Romero et al. (2008), em que os autores definem o processo de mineração de dados educacional em quatro etapas. Os documentos foram lidos na íntegra e agrupados com base na análise de conteúdo a partir da definição de constructos *a priori*, sendo as quatro etapas da MDE: (i) coleta de dados educacionais; (ii) pré-processamento dos dados educacionais; (iii) processamento da mineração de dados; e (iv) interpretação, avaliação e implantação dos resultados. Para esta última etapa, foi utilizada a análise de conteúdo, uma técnica que proporciona apresentar de forma objetiva, sistemática e qualitativa o conteúdo que vai ser analisado (Gil, 2018), por meio da análise por categorias, em que é possível desmembrar os textos que vão ser analisados e reagrupá-los em novas categorias, como forma de identificar similaridades entre os elementos (Bardin, 2011).

5. Resultados

5.1. Análise Descritiva

A presente seção procura apresentar um panorama geral sobre os trabalhos recuperados, de acordo com a metodologia apresentada na seção anterior. Os 29 trabalhos selecionados apresentam, de acordo com a Figura 2, uma oscilação durante os anos analisados. Observa-se que, em 2011, as publicações possuem um pico de 8, e que, na maioria dos anos, se mantêm entre 2 e 3.

Figura 2 - Quantidade de trabalhos por ano



Fonte: Elaborado pelos autores.

De acordo com o Quadro 1, pode-se verificar os 3 autores que mais publicaram sobre o tema. O professor Antonio R. Anaya, do Departamento de Inteligência Artificial da Universidade Nacional de Educação a Distância, sediada em Madri, na Espanha, teve 4 publicações. Ainda, verifica-se que Sebastian Ventura, com 4 publicações, e Cristobal Romero, com 3 publicações, são coautores que se relacionam no desenvolvimento de pesquisas, ambos professores do Departamento de Ciências da Computação da Universidade de Córdoba, na Espanha.

Quadro 1 - Autores que mais publicaram na área

Autores	Artigo	Periódico	Ano
Antonio R. Anaya	Application of machine learning techniques to analyse student interactions and improve the collaboration process	Expert Systems with Applications	2011
	Content-free collaborative learning modeling using data mining	User Modeling and User-Adapted Interaction	2011
	Recommender system in collaborative learning environment using an influence diagram	Expert Systems with Applications	2013
	A visual recommender tool in a collaborative learning experience	Expert Systems with Applications	2016
Sebastian Ventura	Applying Web usage mining for personalizing hyperlinks in Web-based adaptive educational systems	Computers e Education	2009
	An architecture for making recommendations to courseware authors using association rule mining and collaborative filtering	User Modeling and User-Adapted Interaction	2009
	A collaborative educational association rule mining tool	Internet and Higher Education	2011
	Multi-instance genetic programming for predicting student performance in web based educational environments	Applied Soft Computing	2012
Cristobal Romero	Applying Web usage mining for personalizing hyperlinks in Web-based adaptive educational systems	Computers e Education	2009
	An architecture for making recommendations to courseware authors using association rule mining and collaborative filtering	User Modeling and User-Adapted Interaction	2009
	A collaborative educational association rule mining tool	Internet and Higher Education	2011

Fonte: Elaborado pelos autores.

Com relação à origem dos 29 trabalhos selecionados para análise, verifica-se que 12 (41,40%) correspondem à Espanha, 7 (24,13%) a Taiwan, 2 (6,90%) à Inglaterra, à Irlanda e à China. Ainda, verifica-se na amostra outros 8 países com apenas um trabalho relacionado, como Áustria, Grécia, Índia, Holanda, Itália, Sérvia, Malásia e Lituânia. Estes dados mostram que as pesquisas da área possuem grande concentração em determinadas regiões, como identificado pelos achados sobre Espanha e Taiwan.

Quanto às áreas de pesquisa, dado que cada trabalho pode ser classificado em mais de uma área, verifica-se que 82,76% dos artigos correspondem à Ciência da Computação, 27,59% à Engenharia, 24,14% à Educação e Pesquisa Educacional e à Pesquisa de Operações e Ciência do Gerenciamento. Entretanto, outras áreas aparecem com a contagem de apenas um registro, como Negócios e Economia, Psicologia, Telecomunicação, Outros Tópicos em Ciências da Informação e Controle de Sistemas e Automação.

Com relação aos periódicos que mais publicaram sobre o tema estudado (Tabela 1), encontra-se o periódico *Expert Systems with Applications*, que tem como foco publicar artigos que abordam sobre projeto, desenvolvimento, teste, implementação e/ou gerenciamento de sistemas especialistas e inteligentes. Já a revista *Computers e Education* visa a aumentar o conhecimento e a compreensão de como a tecnologia digital pode melhorar a educação. Por fim, o periódico *User Modeling and User-Adapted Interaction* publica pesquisas inovadoras sobre sistemas de computadores interativos que podem ser adaptados a seus usuários e sobre o papel dos modelos de usuário no processo de adaptação. Verifica-se que o escopo de publicação das revistas está relacionado com o tema pesquisado, motivo de serem as que mais publicam sobre a temática.

Tabela 1 - Revistas que mais publicaram sobre o tema

Periódicos	Frequência	%
Expert Systems with Applications	7	21,85
Computers e Education	4	12,50
User Modeling and User-Adapted Interaction	4	12,50

Fonte: Elaborada pelos autores.

No Quadro 2 são apresentados os 10 trabalhos mais citados de acordo com a base ISI Web of Science®, que correspondem a 65,70% de todas as citações dos trabalhos analisados. Como pode ser verificado, o trabalho de Romero, Ventura, Zafra e De Bra (2009) foi o mais relevante e citado por outros estudos, sendo referenciado 75 vezes e correspondendo a 12,86% de todas as citações recebidas pelos trabalhos. Ainda, pode ser verificado que, dentre os trabalhos mais citados, 4 referem-se ao periódico

Computers and Education, 2 ao *Expert Systems with Applications*, e 2 ao *User Modeling and User-Adapted Interaction*, fato que corrobora o destaque das revistas no âmbito pesquisado.

Quadro 2 - Relação dos artigos mais citados na revisão

Qnt	Artigos mais citados
75	Romero, C., Ventura, S., Zafra, A., & De Bra, P. (2009). Applying Web usage mining for personalizing hyperlinks in Web-based adaptive educational systems. <i>Computers and Education</i> , 53(3), 828-840.
48	García, E., Romero, C., Ventura, S., & De Castro, C. (2009). An architecture for making recommendations to courseware authors using association rule mining and collaborative filtering. <i>User Modeling and User-Adapted Interaction</i> , 19(1-2), 99-132.
46	Lin, C. F., Yeh, Y. C., Hung, Y. H., & Chang, R. I. (2013). Data mining for providing a personalized learning path in creativity: An application of decision trees. <i>Computers and Education</i> , 68, 199-210.
46	Lee, C. H., Lee, G. G., & Leu, Y. (2009). Application of automatically constructed concept map of learning to conceptual diagnosis of e-learning. <i>Expert Systems with Applications</i> , 36(2), 1675-1684.
36	Chen, C. M., & Chen, M. C. (2009). Mobile formative assessment tool based on data mining techniques for supporting web-based learning. <i>Computers and Education</i> , 52(1), 256-273.
30	Lara, J. A., Lizcano, D., Martínez, M. A., Pazos, J., & Riera, T. (2014). A system for knowledge discovery in e-learning environments within the European Higher Education Area—Application to student data from Open University of Madrid, UDIMA. <i>Computers and Education</i> , 72, 23-36.
30	García, E., Romero, C., Ventura, S., & De Castro, C. (2011). A collaborative educational association rule mining tool. <i>The Internet and Higher Education</i> , 14(2), 77-88.
25	Aher, S. B., & Lobo, L. M. R. J. (2013). Combination of machine learning algorithms for recommendation of courses in E-Learning System based on historical data. <i>Knowledge-Based Systems</i> , 51, 1-14.
25	Coccea, M., & Weibelzahl, S. (2009). Log file analysis for disengagement detection in e-Learning environments. <i>User Modeling and User-Adapted Interaction</i> , 19(4), 341-385.
22	Wang, Y. H., & Liao, H. C. (2011). Data mining for adaptive learning in a TESL-based e-learning system. <i>Expert Systems with Applications</i> , 38(6), 6480-6485.

Fonte: Dados da pesquisa.

5.2. Análise Semântica

Enfatiza-se que as quatro etapas explicitadas por Romero et al. (2008) servirão como categorias de análise dos resultados expostos a seguir.

5.2.1. Coletar os dados

Com o aumento das tecnologias da informação no campo da educação, os sistemas de gerenciamento foram desenvolvidos e amplamente adotados para armazenar uma ampla gama de dados, incluindo características dos alunos, histórias de aprendizado, notas e outros tipos de informações (Huang, Huang & Chuang, 2016). Sistemas de gerenciamento on-line incluem repositórios de dados educacionais, que armazenam as histórias de aprendizado e os registros dos alunos, de modo que os dados são tesouros de experiências de aprendizado, incluindo o conteúdo visualizado, o envio de tarefas, o tempo gasto e os resultados do questionário, que podem ser analisados para melhorar o desempenho dos alunos (Cuéllar, Delgado & Pegalajar, 2011; Zafra & Ventura, 2012; Lara et al., 2014; Huang et al., 2016).

Como enfatizam Romero et al. (2008), é por meio dos sistemas de gerenciamento, utilizados pelos alunos, que são armazenadas as informações de uso num banco de dados, necessitando ser coletadas. Para isto, foi realizada uma análise dos principais sistemas e eles estão expressos no Quadro 3.

Quadro 3 - Sistemas encontrados na literatura

Sistemas	Descrição	Quem cita
MOODLE	O ambiente modular de aprendizagem orientada a objetos é um sistema de gerenciamento de aprendizado gratuito, flexível, dinâmico e evolutivo, de código aberto, desenvolvido por programadores de todo o mundo, que permite a criação de cursos on-line (Romero et al., 2008; Lara et al., 2014).	Hsieh e Wang (2010); Zafra e Ventura (2012); Aher e Lobo (2013); Lara et al. (2014)
AWBES	Os sistemas educacionais adaptáveis baseados na Web fornecem uma alternativa à tradicional abordagem "just-put-it-on-the-web" (Romero et al., 2009), construindo um modelo dos objetivos, preferências e conhecimentos de cada aluno e usando-o ao longo da interação com este, a fim de se adaptar às suas necessidades (Romero et al., 2009).	Romero et al. (2009)
HTML-Tutor	É um ambiente interativo de aprendizado na web, baseado no NetCoac, que ensina como usar HTML (Coccea & Weibelzahl, 2009).	Coccea e Weibelzahl (2009; 2011)
dotLRN	Apoia os alunos com um espaço de <i>e-learning</i> , tanto um espaço geral, em que os alunos podem encontrar toda a documentação de ensino e serviços para completar as tarefas e comunicar-se com seus pares usando os fóruns, quanto subespaços para trabalhos em equipe (Anaya, Luque & Peinado, 2016)	Anaya, Luque e Peinado (2016)

Sakai	É um sistema de gerenciamento de aprendizado de código aberto desenvolvido pelas principais instituições educacionais do mundo, permitindo uma poderosa colaboração de ensino, aprendizado e pesquisa em uma plataforma corporativa rica em recursos (Acampora et al., 2011).	Acampora et al. (2011).
iHelp	É um sistema de aprendizagem baseado na <i>web</i> que inclui dois aplicativos projetados para dar suporte a alunos e instrutores durante todo o processo de aprendizado: o sistema de discussão iHelp e o sistema de gerenciamento de conteúdo de aprendizagem iHelp (Cocea & Weibelzahl, 2009).	Cocea e Weibelzahl (2009; 2011)
Sistema AHA!	É um sistema de hipermídia adaptativa de código aberto, desenvolvido para suportar um curso on-line com orientação do usuário por meio de explicações condicionais (extras) e ocultação de <i>links</i> condicionais. Oferece instalações de baixo nível para criar a aparência exata para cada aplicação, além de recursos de alto nível para criar a estrutura conceitual de um aplicativo (Romero et al., 2009).	Romero et al. (2009)
Video on Demand (VoD)	O sistema de vídeo sob demanda acomoda cursos de vídeo e é usado em muitas plataformas para fornecer serviços de aprendizado convenientes, suportando práticas repetidas e disponibilizando milhares de horas de conteúdo, permitindo que os alunos aprendam conforme necessário (Xie et al., 2017).	Xie, Zheng e Zhang (2017)
Personalized E-learning System (PELS)	Baseado na teoria da resposta ao item, foi apresentado para serviços de aprendizagem adaptativa individual, reunindo automaticamente os portfólios de aprendizagem úteis de alunos individuais para a avaliação do desempenho durante os processos de aprendizagem (Chen & Chen, 2009).	Chen e Chen (2009)
Personalized Creativity Learning System (PCLS)	Baseado em um sistema multiagente, é composto de duas partes: mecanismos de aprendizagem personalizados e treinamento de criatividade, que têm como objetivo fornecer ao aluno um caminho de aprendizado adaptativo, baseado na faculdade principal, autopercepção, criatividade, estilos de aprendizagem, caminhos de aprendizado e dados sobre criatividade (Lin et al., 2013).	Lin et al. (2013)
AI-test E-learning	É um sistema de aprendizagem adaptativa no ensino de inglês como segunda língua, que otimiza os resultados de aprendizagem dos alunos considerando suas diferentes características e explora o desempenho de aprendizagem de vários alunos usando uma técnica de rede neural artificial (RNA) (Wang & Liao, 2011).	Wang e Liao (2011)
BlackBoard Vista	É utilizado para produzir relatos de materiais de aprendizagem de forma flexível, bem como para possibilitar que os alunos participem de discussões e conversas sincrônicas, criem blogs, revisem vídeos das palestras, usem <i>e-mail</i> etc., além de rastrear atividades dos alunos e interpretá-las (Preidys & Sakalauskas, 2010).	Preidys e Sakalauskas (2010)

Fonte: Dados da pesquisa.

5.2.2. Pré-processamento dos dados

Romero et al. (2008) enfatizam que, quando se utiliza um SGC, a etapa de pré-processamento de dados é um pouco mais simples, dado que emprega autenticação de usuário (proteção por senha), com entradas e saídas identificadas por meio de *login*. Porém, tal etapa é uma fase crucial antes da tarefa analítica, visto que fornece um pré-processamento de dados confiável, sendo importante para se obter dados de qualidade (Mohamad, Ahmad & Sulaiman, 2017). Segundo Romero et al. (2008), são necessárias algumas tarefas em tal etapa, como: (i) selecionar dados, pois é necessário escolher em quais cursos e/ou atividades a mineração pode ser útil; (ii) criar tabelas de resumo, em que possam ser resumidas as informações no nível requerido; (iii) transformar dados em valores numéricos, o que pode ser necessário para aumentar a interpretação e a compreensão; e (iv) transformar os dados no formato necessário do algoritmo ou da estrutura de mineração de dados.

O único trabalho que abordou esta temática foi o de Mohamad et al. (2017), pois, como os próprios autores enfatizam, raramente se discutem os procedimentos de pré-processamento para dados de *e-learning*, o que torna esta etapa uma lacuna de pesquisa para futuros estudos da área. Como limitação do próprio estudo, os autores enfatizam que analisaram apenas os recursos necessários para a parte inicial do procedimento de *e-learning*, como a remoção de dados duplicados e a seleção de dados atuais. Mohamad et al. (2017) propuseram que futuros trabalhos devem aprofundar-se em outros aspectos, como a seleção de dados corretos, a projeção de etapas certas para redução de dados e a construção de recursos educacionais para o pré-processamento.

5.2.3. Aplicar mineração de dados

Atualmente, as ferramentas de mineração de dados são muito complexas, dado que são projetadas

mais para a eficácia e a flexibilidade do que para a simplicidade, sendo necessária a intermediação do administrador do SGC na aplicação de técnicas de mineração de dados, como forma de produzir conteúdo necessário para que tutores consigam tomar decisões quanto ao aprendizado do aluno e aos cursos on-line (Romero et al., 2008).

Para tanto, diante da grande quantidade de dados registrados e armazenados nos repositórios de *e-learning*, é possível aplicar técnicas de mineração de dados para vislumbrar o desempenho de aprendizado dos alunos (Huang et al., 2016). Os pesquisadores de MDE desenvolveram muitas ferramentas para ajudar usuários — como professores, instrutores, universidades, desenvolvedores de cursos, provedores de aprendizagem, entre outros — a resolver diferentes problemas e objetivos educacionais (García et al., 2011).

De acordo com o encontrado na pesquisa, alguns dos principais sistemas de mineração de dados usados para analisar dados de diferentes domínios são apresentados no Quadro 4.

Quadro 4 - Principais sistemas de mineração de dados encontrados

Sistema	Descrição	Quem cita
Sistema Weka	Weka é uma coleção de algoritmos de aprendizado de máquina para tarefas de MD que podem ser aplicados diretamente a um conjunto de dados ou usados a partir do seu próprio código Java. É utilizado para processamento, classificação, regressão, agrupamento, associação e visualização de dados (Romero et al., 2008; Aher & Lobo, 2013).	Preidys e Sakalauskas (2010); Anaya e Boticario (2011a); Köck e Paramythis (2011); Radosav, Brtka e Brtka (2012); Aher e Lobo (2013).
AHA!	É um Applet Java, em que o autor do curso pode executá-lo quando informações suficientes de novos alunos forem coletadas. A interface da ferramenta de mineração é simples, fácil de usar e especificamente orientada para descobrir padrões sequenciais e recomendar <i>links</i> personalizados (Romero et al., 2009).	Romero et al. (2009).
ICDS	Utiliza a combinação do algoritmo de Apriori para fornecer aos professores mapas conceituais construídos pelos alunos, permitindo que os professores diagnostiquem as barreiras de aprendizado e o equívoco dos alunos instantaneamente (Lee et al., 2009).	Lee, Lee e Leu (2009)
CIECoF	É um sistema de recomendação colaborativa aplicado à educação, que ajuda professores a melhorar e manter continuamente cursos de <i>e-learning</i> adaptativos, usando o sistema de <i>feedback</i> . O sistema é baseado na arquitetura cliente-servidor, que aplica um algoritmo de associação localmente nos dados dos alunos em um curso on-line (García et al., 2009).	García et al. (2009); García et al. (2011).

Fonte: Dados da pesquisa.

5.2.4. Interpretar, avaliar e implantar os resultados

Dentro de uma instituição educacional baseada no ensino a distância, existem muitos grupos de interesse quanto às informações obtidas pela MDE, de modo que podem utilizar as informações obtidas em diversos âmbitos (Kotsiantis, 2012). Apresenta-se a seguir (Quadro 5) uma sistematização dos resultados de implementação observados nos artigos selecionados, como forma de explicitar na prática como a MDE pode corroborar no aperfeiçoamento da gestão do *e-learning*.

As categorias foram obtidas a partir da leitura dos trabalhos selecionados e agrupadas por similaridade, obtendo-se as seguintes: (i) monitoramento e aperfeiçoamento de práticas de aprendizagem: uma das principais categorias, dado que descobrir padrões de aprendizagem dos alunos e acompanhá-los é importante para que o aprendizado seja mais significativo e personalizado; (ii) técnicas de recomendação, em que dados de uso dos alunos podem fornecer informação suficiente para que os sistemas de aprendizagem recomendem possíveis materiais complementares, *links*, cursos etc.; (iii) desengajamento e desistência, de modo que é possível identificar alunos desinteressados pela forma que utilizam os sistemas de *e-learning*, possibilitando atuar sobre estes alunos, motivando-os; (iv) previsão de desempenho dos alunos, de forma a identificar previamente possíveis alunos com baixo rendimento a partir da sua utilização do sistema ou do desempenho de alunos com características similares, proporcionando interferir no seu aprendizado; (v) colaboração, uma vez que um dos principais problemas é desenvolver atividades em grupo, estando a distância; (vi) melhoramento da arquitetura de sistemas de *e-learning*, como forma de melhorar e facilitar o uso de sistemas menos complexos, mas eficientes, por professores e gestores.

Quadro 5 – Sistematização dos resultados de implementação

Implementação	Descrição	Autores
Monitoramento e aperfeiçoamento de práticas de aprendizagem	Dados os diferentes métodos de aprendizagem dos alunos, a MDE auxilia na análise dos fatores e dos estilos de aprendizagem, preparando materiais de aprendizagem individualizados e melhorando a apresentação do curso, a qualidade dos estudos e dos resultados.	Preidys e Sakalauskas (2010)
	Metodologias de MDE podem fornecer ferramentas para analisar as necessidades e preferências do aluno e, conseqüentemente, gerar experiências de aprendizado eficientes, personalizadas e flexíveis.	Acampora et al. (2011)
	O monitoramento e a interpretação das atividades sequenciais dos alunos têm o potencial de melhorar a adaptabilidade e a personalização nos ambientes educacionais, resultando na descoberta de novas informações semanticamente significativas sobre os alunos.	Köck e Paramythis (2011)
	Com base no gênero, nos tipos de personalidade e nos níveis de ansiedade, a MDE pode definir diferentes níveis de conteúdo para alunos com diferentes características, permitindo um aprendizado personalizado.	Wang e Liao (2011)
	A personalização de um ambiente de <i>e-learning</i> para otimizar o aprendizado é importante, como, por exemplo, o desenvolvimento de um sistema de aprendizado para fornecer caminhos personalizados para otimizar o desempenho da criatividade.	Lin et al. (2013)
	Um dos desafios do <i>e-learning</i> é a lacuna espacial e temporal entre o professor e o acompanhamento do aluno, de modo que a MDE pode extrair informações úteis para os professores sobre a interação do aluno com o ambiente de aprendizagem.	Lara et al. (2014)
	Por meio da MDE é possível descobrir padrões de aprendizagem dos alunos, de modo a identificar possíveis similaridades de comportamento entre alunos com características semelhantes e agir sobre eles.	Huang et al. (2016)
	Muitos mecanismos de avaliação não são eficazes para o desempenho da aprendizagem nos sistemas de <i>e-learning</i> , principalmente a avaliação somativa tradicional, que considera apenas resultados finais e não se preocupa com o processo de aprendizagem. Novos métodos de avaliação do desempenho da aprendizagem devem avaliar o que os alunos aprenderam durante o processo.	Chen e Chen (2009)
	Métodos como o mapa conceitual, com base em MDE, devem ser implementados para retratar a estrutura do conhecimento e diagnosticar os conceitos errôneos e as barreiras de aprendizagem dos alunos no <i>e-learning</i> de maneira instantânea.	Lee et al. (2009)
Técnicas de recomendação	Por meio da MDE, é possível recomendar cursos a um aluno com base na escolha de outros alunos, ou seja, alunos que fazem um mesmo curso podem ter o mesmo interesse por cursos parecidos no futuro.	Aher e Lobo (2013)
	A MDE possibilita descobrir informações importantes através dos dados de uso dos alunos na forma de regras de recomendação, juntamente com a possibilidade de compartilhar as regras com professores e especialistas com perfis parecidos.	García et al. (2009; 2011)
	A MDE possibilita o desenvolvimento de um sistema de recomendação que utiliza a mineração de uso da <i>Web</i> para recomendar os <i>links</i> a serem visitados em um sistema educacional adaptável baseado na <i>Web</i> .	Romero et al. (2009)
	Por meio da MDE, é possível desenvolver um sistema de suporte à aprendizagem que utiliza recomendação de objetos ou de documentos adequados para cada unidade dos cursos, a fim de facilitar um aprendizado mais eficiente.	Hsieh e Wang (2010)
Desengajamento e desistência	A MDE auxilia na predição de um usuário se retirar prematuramente e permanentemente de um curso, intervindo com estratégias para prolongar o tempo de uso do aluno, recomendando-lhe materiais personalizados.	Xie et al. (2017)
	Existem diferentes tipos de alunos descomprometidos, os que passam as páginas sem lê-las e os que passam muito tempo em uma página, de modo que, por meio da MDE, pode-se identificá-los e intervir sobre eles, promovendo alterações no sistema de <i>e-learning</i> como forma de engajá-los.	Cocca e Weibelzahl (2009; 2011)
Previsão de desempenho do aluno	Prever o desempenho dos alunos é essencial, principalmente para identificar os que têm baixa probabilidade de passar no curso e que precisam de uma maior atenção, para que os tutores possam destinar mais tempo e concentração a eles.	Zafra e Ventura (2012)
	As principais características demográficas e as notas em um pequeno número de tarefas dos alunos podem, a partir de técnicas de MDE, prever o seu desempenho e suas notas, auxiliando a ação de professores e tutores.	Kotsiantis (2012)
	Um sistema de alerta precoce pode ajudar a identificar alunos em risco ou prever o desempenho do aprendizado, analisando portfólios registrados em um sistema de gerenciamento.	Hu et al. (2014)

	A MDE possibilita identificar antecipadamente as ações dos alunos, detectando situações em que cometem erros ou ficam presos numa atividade, fornecendo previsões razoavelmente boas e suporte de <i>feedback</i> mais adaptado.	Luzcando, Ramirez e Lobo (2017)
Colaboração	Em ambientes de <i>e-learning</i> que usam a estratégia de colaboração, fornecer aos participantes um conjunto de serviços de comunicação pode não ser suficiente para garantir o aprendizado colaborativo. Portanto, é necessário analisar a colaboração regularmente e com frequência. A MDE pode auxiliar a monitorar e a desenvolver métodos e modelos de colaboração entre os alunos com base na análise dos dados de interação entre eles, incorporando um componente social na educação a fim de verificar se a colaboração ocorreu e propor modelos mais enfáticos de interação sem a intervenção humana.	Anaya e Boticario (2011a; 2011b); Anaya, Luque e García-Saiz (2013); Anaya et al. (2016)
Melhoramento da arquitetura de sistemas de <i>e-learning</i>	Na maioria das vezes, a MDE extrai o conhecimento temporal com base principalmente na frequência de eventos, ignorando sua duração. Os autores propõem um método que extrai grandes dados de aprendizado, mais completo, levando em consideração tanto a frequência quanto a duração.	Xie, Zheng e Zhang (2018)
	Desenvolver novos modelos de arquitetura com a capacidade de melhor analisar os sistemas de <i>e-learning</i> é essencial, principalmente a partir de novas regressões para analisar as redes neurais de aprendizado incremental desses sistemas.	Radosav et al. (2012)
	Este estudo apresentou o processo de pré-processamento de dados, uma fase crucial antes da implementação da tarefa analítica, para um conjunto de dados, desde a coleta até a sua transformação e algumas regras.	Mohamad et al. (2017)
	Formulação e interpretação de plataformas de <i>e-learning</i> como redes sociais, desenvolvendo uma rede social personalizada. Contribui para que alunos, professores e recursos de aprendizagem tenham uma melhor compreensão de sua estrutura social e, portanto, tomem ou melhorem decisões sobre o processo.	Cuéllar et al. (2011)

Fonte: Dados da pesquisa.

6. Conclusão

O estudo teve como objetivo analisar o que tem sido publicado sobre a mineração de dados educacionais em contextos de *e-learning*, como forma de colaborar na construção do seu conhecimento, através de uma revisão sistemática de literatura. Para tanto, foi realizada uma análise descritiva sobre o tema abordado, caracterizando os principais achados e, em seguida, realizando uma análise semântica sobre os principais temas encontrados.

Por meio da análise descritiva foi possível perceber que o tema não possui nem grandes quantidades de publicações ao longo dos anos, nem autores e revistas que se destacam com muitas publicações, o que demonstra que a pesquisa na área ainda não está consolidada e pode possuir diversas lacunas. O tema ainda está na sua infância, necessitando de mais pesquisas na área sob diversas perspectivas (Lee, 2017; Farid et al., 2018). Ainda, com relação aos trabalhos mais citados, verifica-se que eles não possuem expressividade de citações, com poucos artigos sendo referenciados mais de 20 vezes.

Quanto à análise semântica, ela foi baseada no trabalho de Romero et al. (2008), que definem o processo de mineração de dados educacional em quatro etapas: (i) coleta de dados educacionais; (ii) pré-processamento dos dados educacionais; (iii) processamento da mineração de dados; e (iv) interpretação, avaliação e implantação dos resultados.

Quanto à etapa da coleta de dados, identificaram-se diversos sistemas de gerenciamento de cursos em que os dados ficam armazenados e precisam ser coletados para posteriormente serem analisados. Para tanto, sistemas mais conhecidos como MOODLE, além de outros mais específicos e menos disseminados, foram utilizados por diversos estudos como forma de analisar seus dados armazenados. O conhecimento arquivado em grandes bancos de dados pode ser útil para diversos tipos de atores, como professores, alunos, dentre outros, de forma que identificá-los e coletá-los torna-se uma etapa fundamental no processo de MDE.

A etapa de pré-processamento de dados não apresentou trabalhos relacionados a esta perspectiva, demonstrando uma possível lacuna para futuras pesquisas, dado que, como apresentado, é uma importante etapa do processo de MDE, por ser o momento em que se lapidam os dados para que possam ser analisados nas próximas etapas. Na terceira etapa, processamento de dados, apresentaram-se alguns dos principais sistemas de mineração de dados utilizados pelos estudos como forma de analisar e interpretar os dados coletados. Enfatiza-se ainda que foram abordados apenas os sistemas de mineração de dados, não sendo analisadas as técnicas, nem as abordagens e algoritmos utilizados pelos artigos, constituindo uma sugestão para pesquisas futuras.

A última parte da mineração de dados apresentou os resultados encontrados e implementados

pelas pesquisas analisadas. Como pode ser verificado, o processo de MDE é fundamental no contexto de gestão do *e-learning*, dado que existem diversos atores interessados nas análises dos dados. Assim, uma boa execução da MDE pode auxiliar na avaliação e no melhoramento de diversos elementos de um sistema de *e-learning*, aperfeiçoando os cursos on-line e o processo de ensino-aprendizado a distância.

O estudo contribui para aspectos práticos e teóricos, uma vez que, ao condensar as etapas da perspectiva de MDE na gestão do *e-learning*, pode-se corroborar a teoria, demonstrando o que vem sendo pesquisado na área. Quanto aos aspectos práticos, o estudo colabora com a apresentação do passo a passo do processo de MDE, assim como os sistemas, práticas, técnicas e outras questões relacionadas ao tema que existem e podem ser utilizados para o aperfeiçoamento da MDE para interessados, como escolas, professores e alunos.

Por fim, este estudo, apesar de apresentar avanços, tem algumas limitações. Dado que é uma revisão sistemática, o procedimento metodológico utilizado pode ocasionar a não inclusão de possíveis documentos, assim como a exclusão de documentos que poderiam contribuir para os resultados. Ainda, apesar de a base de busca utilizada, Web of Science, ser uma das mais completas bases de pesquisa científica existentes, tornando-a um dos bancos de dados globais de citações mais confiáveis do mundo, com mais de 1,7 bilhão de referências com informações multidisciplinares (Clarivate Analytics, 2020), a utilização de apenas uma base de dados para a coleta dos artigos da revisão constitui outra limitação da pesquisa. Futuros estudos podem utilizar outras estratégias de busca além das apresentadas no presente artigo, bem como utilizar outras bases de dados como forma de coletar os artigos a serem analisados.

Referências

- Acampora, G., Cadenas, J. M., Loia, V., & Ballester, E. M. (2011). A multi-agent memetic system for human-based knowledge selection. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 41(5), 946-960.
- Agrusti, F., Bonavolontà, G., & Mezzini, M. (2019). University Dropout Prediction through Educational Data Mining Techniques: A Systematic Review. *Journal of e-Learning and Knowledge Society*, 15(3), 161-182.
- Aher, S. B., & Lobo, L. M. R. J. (2013). Combination of machine learning algorithms for recommendation of courses in E-Learning System based on historical data. *Knowledge-Based Systems*, 51, 1-14.
- Aldowah, H., Al-Samarraie, H., & Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics*, 37, 13-49.
- Anaya, A. R., & Boticario, J. G. (2011a). Application of machine learning techniques to analyse student interactions and improve the collaboration process. *Expert Systems with Applications*, 38(2), 1171-1181.
- Anaya, A. R., & Boticario, J. G. (2011b). Content-free collaborative learning modeling using data mining. *User Modeling and User-Adapted Interaction*, 21(1-2), 181-216.
- Anaya, A. R., Luque, M., & García-Saiz, T. (2013). Recommender system in collaborative learning environment using an influence diagram. *Expert Systems with Applications*, 40(18), 7193-7202.
- Anaya, A. R., Luque, M., & Peinado, M. (2016). A visual recommender tool in a collaborative learning experience. *Expert Systems with Applications*, 45, 248-259.
- Angeli, C., Howard, S. K., Ma, J., Yang, J., & Kirschner, P. A. (2017). Data mining in educational technology classroom research: Can it make a contribution?. *Computers & Education*, 113, 226-242.
- Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers & Education*, 113, 177-194.
- Bakhshinategh, B., Zaiane, O. R., ElAtia, S., & Ipperciel, D. (2018). Educational data mining applications and tasks: A survey of the last 10 years. *Education and Information Technologies*, 23(1), 537-553.
- Bardin, L. (2011). *Análise de conteúdo*. 1. São Paulo: Edições, 70, 223.
- Chalaris, M., Gritzalis, S., Maragoudakis, M., Sgouropoulou, C., & Tsolakidis, A. (2014). Improving quality of educational processes providing new knowledge using data mining techniques. *Procedia-Social and Behavioral Sciences*, 147, 390-397.
- Chamizo-Gonzalez, J., Cano-Montero, E. I., Urquia-Grande, E., & Muñoz-Colomina, C. I. (2015). Educational data mining for improving learning outcomes in teaching accounting within higher education. *The International Journal of Information and Learning Technology*, 32(5), 272-285.
- Chen, C. M., & Chen, M. C. (2009). Mobile formative assessment tool based on data mining techniques for

- supporting web-based learning. *Computers & Education*, 52(1), 256-273.
- Clarivate Analytics. (2020, May 31) *Web of Science™: Confident research begins here*. [Web page]. Retrieved from <https://clarivate.com/webofsciencegroup/solutions/web-of-science/>
- Cocea, M., & Weibelzahl, S. (2009). Log file analysis for disengagement detection in e-Learning environments. *User Modeling and User-Adapted Interaction*, 19(4), 341-385.
- Cocea, M., & Weibelzahl, S. (2011). Disengagement detection in on-line learning: Validation studies and perspectives. *IEEE transactions on learning technologies*, 4(2), 114-124.
- Costa, E. B., Fonseca, B., Santana, M. A., de Araújo, F. F., & Rego, J. (2017). Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. *Computers in Human Behavior*, 73, 247-256.
- Cuéllar, M. P., Delgado, M., & Pegalajar, M. C. (2011). Improving learning management through semantic web and social networks in e-learning environments. *Expert Systems with Applications*, 38(4), 4181-4189.
- Ćukušić, M., Alfirević, N., Granić, A., & Garača, Ž. (2010). e-Learning process management and the e-learning performance: Results of a European empirical study. *Computers & Education*, 55(2), 554-565.
- Dutt, A., Ismail, M. A., & Herawan, T. (2017). A systematic review on educational data mining. *Ieee Access*, 5, 15991-16005.
- Farid, S., Ahmad, R., Alam, M., Akbar, A., & Chang, V. (2018). A sustainable quality assessment model for the information delivery in E-learning systems. *Information Discovery and Delivery*, 46(1), 1-25.
- Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R., & Van Erven, G. (2019). Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil. *Journal of Business Research*, 94, 335-343.
- Galvão, C. M., Sawada, N. O., & Trevizan, M. A. (2004). Revisão sistemática. *Rev Latino-am enfermagem*, 12(3), 549-56.
- García, E., Romero, C., Ventura, S., & De Castro, C. (2009). An architecture for making recommendations to courseware authors using association rule mining and collaborative filtering. *User Modeling and User-Adapted Interaction*, 19(1-2), 99-132.
- García, E., Romero, C., Ventura, S., & De Castro, C. (2011). A collaborative educational association rule mining tool. *The Internet and Higher Education*, 14(2), 77-88.
- Gil, A. C. (2018). *Como elaborar projetos de pesquisa*. 6. São Paulo: Atlas.
- Gregg, D. G. (2007). E-learning agents. *The Learning Organization*, 14(4), 300-312.
- Haq, A. U., Magoulas, G., Jamal, A., Majeed, A., & Sloan, D. (2018). Users' perceptions of e-learning environments and services effectiveness: The emergence of the concept functionality model. *Journal of Enterprise Information Management*, 31(1), 89-111.
- Harrati, N., Bouchrika, I., & Mahfouf, Z. (2017). Investigating the uptake of educational systems by academics using the technology to performance chain model. *Library Hi Tech*, 35(4), 629-648.
- Hsieh, T. C., & Wang, T. I. (2010). A mining-based approach on discovering courses pattern for constructing suitable learning path. *Expert systems with applications*, 37(6), 4156-4167.
- Hu, Y. H., Lo, C. L., & Shih, S. P. (2014). Developing early warning systems to predict students' on-line learning performance. *Computers in Human Behavior*, 36, 469-478.
- Huang, T. C. K., Huang, C. H., & Chuang, Y. T. (2016). Change discovery of learning performance in dynamic educational environments. *Telematics and Informatics*, 33(3), 773-792.
- Joorabchi, A., English, M., & Mahdi, A. E. (2016). Text mining stackoverflow: An insight into challenges and subject-related difficulties faced by computer science learners. *Journal of Enterprise Information Management*, 29(2), 255-275.
- Khodabandeh, A., Afshari, H., & Manian, A. (2010, June). Critical factors affecting e-learner's satisfaction an empirical study. In *EdMedia+ Innovate Learning*. Association for the Advancement of Computing in Education (AACE).
- Köck, M., & Paramythis, A. (2011). Activity sequence modelling and dynamic clustering for personalized e-learning. *User Modeling and User-Adapted Interaction*, 21(1-2), 51-97.
- Kotsiantis, S. B. (2012). Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades. *Artificial Intelligence Review*, 37(4), 331-344.
- Lamph, G., Sampson, M., Smith, D., Williamson, G., & Guyers, M. (2018). Can an interactive e-learning training package improve the understanding of personality disorder within mental health professionals?. *The Journal of Mental Health Training, Education and Practice*, 13(2), 124-134.

- Lara, J. A., Lizcano, D., Martínez, M. A., Pazos, J., & Riera, T. (2014). A system for knowledge discovery in e-learning environments within the European Higher Education Area–Application to student data from Open University of Madrid, UDIMA. *Computers & Education*, 72, 23-36.
- Lee, C. H., Lee, G. G., & Leu, Y. (2009). Application of automatically constructed concept map of learning to conceptual diagnosis of e-learning. *Expert Systems with Applications*, 36(2), 1675-1684.
- Lee, Y. J. (2017). Modeling students' problem solving performance in the computer-based mathematics learning environment. *The International Journal of Information and Learning Technology*, 34(5), 385-395.
- Lin, C. F., Yeh, Y. C., Hung, Y. H., & Chang, R. I. (2013). Data mining for providing a personalized learning path in creativity: An application of decision trees. *Computers & Education*, 68, 199-210.
- Liñán, L. C., & Pérez, Á. A. J. (2015). Educational Data Mining and Learning Analytics: differences, similarities, and time evolution. *International Journal of Educational Technology in Higher Education*, 12(3), 98-112.
- Luo, N., Zhang, M., & Qi, D. (2017). Effects of different interactions on students' sense of community in e-learning environment. *Computers & Education*, 115, 153-160.
- Luzcando, D. R., Ramirez, J., & Lobo, M. B. (2017). Predicting student actions in a procedural training environment. *IEEE Transactions on Learning Technologies*, 10(4), 463-474.
- Male, G., & Pattinson, C. (2011). Enhancing the quality of e-learning through mobile technology: A socio-cultural and technology perspective towards quality e-learning applications. *Campus-Wide Information Systems*, 28(5), 331-344.
- Manjarres, A. V., Sandoval, L. G. M., & Suárez, M. S. (2018). Data mining techniques applied in educational environments: Literature review. *Digital Education Review*, (33), 235-266.
- Mohamad, N., Ahmad, N. B., & Sulaiman, S. (2017). Data pre-processing: a case study in predicting student's retention in MOOC. *Journal of Fundamental and Applied Sciences*, 9(4S), 598-613.
- Mohamad, S. K., & Tasir, Z. (2013). Educational data mining: A review. *Procedia-Social and Behavioral Sciences*, 97(2013), 320-324.
- Oliveira, P. C. D., Cunha, C. J. C. D. A., & Nakayama, M. K. (2016). Learning Management Systems (LMS) and e-learning management: an integrative review and research agenda. *JISTEM-Journal of Information Systems and Technology Management*, 13(2), 157-180.
- Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Journal of Educational Technology & Society*, 17(4), 49-64.
- Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert systems with applications*, 41(4), 1432-1462.
- Preidys, S., & Sakalauskas, L. (2010). Analysis of students' study activities in virtual learning environments using data mining methods. *Technological and economic development of economy*, 16(1), 94-108.
- Radosav, D., Brtka, E., & Brtka, V. (2012). Mining Association Rules from Empirical Data in the Domain of Education. *International Journal of Computers Communications & Control*, 7(5), 933-944.
- Rodrigues, M. W., Isotani, S., & Zárate, L. E. (2018). Educational Data Mining: A review of evaluation process in the e-learning. *Telematics and Informatics*, 35(6), 1701-1717.
- Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert systems with applications*, 33(1), 135-146.
- Romero, C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(6), 601-618.
- Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, 51(1), 368-384.
- Romero, C., Ventura, S., Zafra, A., & De Bra, P. (2009). Applying Web usage mining for personalizing hyperlinks in Web-based adaptive educational systems. *Computers & Education*, 53(3), 828-840.
- Shahiri, A. M., & Husain, W. (2015). A review on predicting student's performance using data mining techniques. *Procedia Computer Science*, 72, 414-422.
- Simo, A., Barbulescu, C., & Kilyeni, S. (2015). Current practices in e-learning: A case study for electrical power engineering in higher education. *Procedia-Social and Behavioral Sciences*, 191, 605-610.
- Siqueira, S. W. M., Braz, M. H. L. B., & Melo, R. N. (2007). Modeling e-learning content. *International Journal of Web Information Systems*.

- Slater, S., Joksimović, S., Kovanovic, V., Baker, R. S., & Gasevic, D. (2017). Tools for educational data mining: A review. *Journal of Educational and Behavioral Statistics*, 42(1), 85-106.
- Tan, C. J., Lim, T. Y., Bong, C. W., & Liew, T. K. (2017). A multi-objective evolutionary algorithm-based soft computing model for educational data mining: A distance learning experience. *Asian Association of Open Universities Journal*, 12(1), 106-123.
- Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British journal of management*, 14(3), 207-222.
- Wang, Y. H., & Liao, H. C. (2011). Data mining for adaptive learning in a TESL-based e-learning system. *Expert Systems with Applications*, 38(6), 6480-6485.
- Weiser, O., Blau, I., & Eshet-Alkalai, Y. (2018). How do medium naturalness, teaching-learning interactions and Students' personality traits affect participation in synchronous E-learning?. *The Internet and Higher Education*, 37, 40-51.
- Xie, T., Zheng, Q., & Zhang, W. (2018). Mining temporal characteristics of behaviors from interval events in e-learning. *Information Sciences*, 447, 169-185.
- Xie, T., Zheng, Q., Zhang, W., & Qu, H. (2017). Modeling and predicting the active video-viewing time in a large-scale E-learning system. *IEEE Access*, 5, 11490-11504.
- Zafra, A., & Ventura, S. (2012). Multi-instance genetic programming for predicting student performance in web based educational environments. *Applied Soft Computing*, 12(8), 2693-2706.