

Mapeamento sistemático da literatura brasileira sobre educational data mining e learning analytics

/

Systematic mapping of brazilian literature on educational data mining and learning analytics

DOI:10.34117/bjdv6n1-005

Recebimento dos originais: 30/11/2019

Aceitação para publicação: 02/01/2020

Francisco Alan de Oliveira Santos

Mestrando do Programa de Pós-Graduação em Engenharia da Computação e Sistemas (PECS) da Universidade Estadual do Maranhão – UEMA, Professor de Ensino Básico, Técnico e Tecnológico
 Instituição: Instituto Federal de Educação do Maranhão – IFMA / Campus Coelho Neto
 Endereço: MA-043, s/n – Olho D’Aguinha, CEP 65620-000, Coelho Neto – MA, Brasil
 E-mail: franciscoalan.santos@ifma.edu.br

Mardoqueu Sousa Telvina

Instituição: Instituto Federal de Educação do Maranhão – IFMA / Campus Coelho Neto
 Endereço: MA-043, s/n – Olho D’Aguinha, CEP 65620-000, Coelho Neto – MA, Brasil
 E-mail: mardoqueu.telvina@ifma.edu.br

Márcio Fabiano Oliveira de Moura Santos

Especialista em Docência do Ensino Superior pela Faculdade Internacional Signorelli, Brasil
 Instituição:
 Endereço:
 E-mail: mfomoura@gmail.com

RESUMO

Este artigo apresenta um Mapeamento Sistemático (MS) sobre a aplicação de Educational Data Mining (EDM) e Learning Analytics (LA) em pesquisas realizadas no contexto do Brasil no quadriênio 2014-2017. Foram consultadas as bases de periódicos, conferências e workshops nacionais sobre Informática na Educação (IE), considerando os artigos disponíveis no portal de publicações da Comissão Especial de Informática na Educação (CEIE). A pesquisa teve como objetivo descrever uma visão geral do estado das pesquisas nacionais com EDM e LA no período analisado. Os resultados demonstraram um aumento considerável na exploração desses tópicos, com uma predominância nos artigos provenientes do SBIE.

Palavras-chave: Mineração de Dados Educacionais; Analíticas de Aprendizagem, MDE; AA.

ABSTRACT

This paper presents a Systematic Mapping (MS) on the application of Educational Data Mining (EDM) and Learning Analytics (LA) in research carried out in the context of Brazil in the years from 2014 to 2017. The bases of journals, conferences and national workshops on Computers in Education (CE) were consulted, considering the papers available in the publications portal of the Special Commission on Informatics in Education (CEIE). The research aimed to describe an overview of the state of national research with EDM and LA. The results showed a considerable increase in the exploitation of these topics, with a predominance in articles from the SBIE.

Keywords: Educational Data Mining; Learning Analytics; EDM; LA.

1 INTRODUÇÃO

O avanço das Tecnologias da Informação e Comunicação (TIC) tem contribuído para proporcionar recursos computacionais voltados para a Educação, permitindo o surgimento de abordagens inovadoras e criando um cenário sem precedentes, onde as possibilidades de exploração desses recursos tem sido alvo de pesquisas multidisciplinares envolvendo as áreas de Computação e Educação (GOLDSCHMIDT et al., 2015). O uso massivo dos Ambientes Virtuais de Aprendizagem (VILLARDI et al., 2010), Sistemas Tutores Inteligentes (WENGER, 1987) e Sistemas Hipermídia Adaptativos Educacionais (BRUSILOVSKKY, 2001) permite que sejam coletados diversos dados sobre os seus usuários, gerando registros cumulativos das suas atividades nesses ambientes (GOLDSCHMIDT et al., 2015). Essa grande quantidade de dados é decorrente das diversas possibilidades de interação nessas plataformas, produzindo massivos volumes de dados.

Como consequência, nos últimos anos, dois tópicos de pesquisa têm emergido na área de Informática na Educação (IE) com o propósito de analisar estes dados e utilizar os resultados para melhorar o processo de ensino-aprendizagem (MOISSA et al., 2015), a saber: *Educational Data Mining* (EDM) — em português, Mineração de Dados Educacionais e *Learning Analytics* (LA) — em português, Analíticas de Aprendizagem. De acordo com Moissa et al. (2015) esses dois tópicos de pesquisa têm se confundido no meio acadêmico, não sendo muito claros os limites de aplicação de cada um. Porém ainda segundo Moissa et al. (2015), existem peculiaridades no uso de algoritmos, instrumentos e técnicas em ambos os tópicos.

A EDM é uma área de pesquisa em crescimento (GOLDSCHMIDT et al., 2015). Seu foco principal é o desenvolvimento de métodos para explorar conjuntos de dados coletados em ambientes educacionais e utilizar os resultados para melhorar o processo de ensino-aprendizagem (BAKER; INVENTADO, 2014). Em geral, as aplicações de EDM buscam novos conhecimentos que permitam melhor compreender o processo educacional a partir da transformação de dados brutos em informação útil que possa contribuir positivamente para a prática do processo (GOLDSCHMIDT et al., 2015). Dentre os conjuntos de dados cuja exploração pode produzir a percepção de aspectos pedagógicos importantes, podem-se citar: quantidade de acessos, media de tempo on-line, periodicidade de participação em fóruns de discussão, frequência de entrega de atividades na plataforma, videoaulas assistidas, entre outras.

Já o termo LA, refere-se à aplicação de técnicas de classificação, mensuração e qualificação de um grande volume de dados educacionais (SCLATER, 2017). Essa área tem ganhado força nos últimos anos e sua adoção tem sido feita principalmente na forma de ferramentas de análise automatizada de bancos de dados sobre educação. A mineração desses dados tem causado descobertas

importantes. Dentre as descobertas possíveis em LA podem-se citar: Determinação de risco de evasão de alunos a partir de sua performance no curso ou da combinação desses dados com outras variáveis, como perfil socioeconômico, identificação de alunos que precisam de ajuda, análise de dados atuais e históricos para prever desempenho de estudantes, entre outros.

A partir deste cenário, este estudo tem o objetivo de contribuir com um Mapeamento Sistemático (MS) da literatura brasileira, baseado nos principais periódicos e eventos de IE do Brasil entre 2014 e 2017, considerando os artigos disponíveis no portal de publicações da Comissão Especial de Informática na Educação (CEIE)¹. Dessa forma, com a descrição dos estudos da área, objetiva-se contribuir com a comunidade brasileira de IE, construindo uma base relevante de conhecimento, para que pesquisadores de EDM e LA conheçam o desenvolvimento de trabalhos envolvendo esses tópicos no cenário brasileiro, e direcionem seus esforços a realização de pesquisas que preencham as lacunas nessas áreas.

O restante deste artigo encontra-se organizado da seguinte maneira: A Seção 2 trata dos trabalhos relacionados. A Seção 3 detalha a condução deste MS da literatura. A Seção 4 mostra a análise dos resultados encontrados. Por fim, a Seção 5 apresenta as conclusões acerca deste estudo.

2 TRABALHOS RELACIONADOS

Em Moissa et al. (2014), através de um MS de trabalhos de LA realizados entre 1996 e 2014, pode-se verificar que a grande maioria dos trabalhos de LA está relacionado a diversas técnicas e ao uso de dados navegacionais, embora diferentes tipos de dados sejam analisados. Outro achado desse mapeamento foi que o público-alvo da maioria dos estudos são pesquisadores, com poucos trabalhos tendo professores e alunos como interessados.

Rodrigues et al. (2014) apresentam uma análise dos principais trabalhos de pesquisa na área de EDM publicados no Brasil de 2006 a 2014. Os resultados mostraram que a maioria dos trabalhos foi realizada através da aplicação direta das técnicas de mineração de dados e posterior processamento das informações em plataforma específicas. Grande parte destes trabalhos tratou de análises em bases de dados de sistemas do tipo ambientes virtuais de aprendizagem, predominantemente o *Moodle*.

Moissa et al. (2015), realizaram um MS sobre EDM e LA. A busca por trabalhos foi conduzida nos repositórios: Web of Knowledge², Engineering Village³, Scopus SciVerse⁴, IEEE Xplore⁵, ACM

¹ <https://www.br-ie.org/pub/index.php/index>

² <https://www.webofknowledge.com>

³ <https://www.engineeringvillage.com>

⁴ <https://www.scopus.com>

⁵ <https://www.ieeexplore.ieee.org/>

Digital Library⁶, Science Direct⁷ e Springer Link⁸. O objetivo do trabalho foi diferenciar as áreas de LA e EDM. Diferenças foram encontradas no uso de algoritmos, instrumentos e técnicas.

Portanto, apesar da similaridade temática entre este estudo e os demais mapeamentos mencionados, a diferença se dá no seu objetivo, nas fontes consultadas e no período de tempo considerado, uma vez que foram analisados trabalhos dos anos 2014 a 2017. Uma característica marcante nesta abordagem em relação às demais é por seu enfoque no cenário científico brasileiro, isto é, todas as considerações, procedimentos e questões de pesquisa têm o objetivo de traçar um perfil das pesquisas publicadas no Brasil, assim como em Rodrigues et al. (2014), este último, porém, sendo datado de 2014.

3 MÉTODO DE PESQUISA

MS corresponde a uma investigação ampla envolvendo estudos primários relacionados com um tópico de pesquisa específico, cujo objetivo é identificar e classificar o conteúdo relacionado com o tópico de pesquisa, visando identificar as evidências disponíveis sobre esse tópico (KITCHENHAM; CHARTERS, 2017). Sendo assim, este trabalho se caracteriza como um estudo secundário teórico-conceitual. É importante ressaltar que não está sendo considerada a avaliação da qualidade dos trabalhos.

O processo de MS envolve três fases principais: Planejamento da Revisão, Condução da Revisão e Publicação dos Resultados (KITCHENHAM; CHARTERS, 2007). Essas fases, bem como suas atividades, são conduzidas de modo iterativo. Para realizar este MS, foi utilizada como referência a metodologia proposta em Keele (2007). Tal escolha se justifica pela sua adequação à área da Computação. Desta forma utilizou-se o protocolo sugerido resultando nas atividades de: 1. Definição das questões de pesquisa; 2. Definição da estratégia de pesquisa; 3. Criação dos critérios de seleção e 4. Estratégia para extração dos dados relevantes.

3.1 QUESTÕES DE PESQUISA

Definiu-se como objetivo desta pesquisa identificar os elementos utilizados ou referenciados na utilização de EDM e LA no contexto brasileiro, considerando as publicações do CEIE. Desta forma, este MS tem por objetivo responder à seguinte pergunta: “Quais são as principais abordagens envolvendo EDM e LA investigadas no domínio brasileiro no período analisado?”. Para isso, foram definidas as questões de pesquisa, cuja elaboração foi baseada nos cinco aspectos presentes na maioria dos MS, apontados por Petersen et al. (2015), a saber: tipos de contribuição, tipos de pesquisa realizados, foco de estudo, veículo de publicação e método de pesquisa.

⁶ <http://dl.acm.org>

⁷ <https://www.sciencedirect.com>

⁸ <https://link.springer.com>

De acordo com Felizardo et al. (2017), um tipo de questionamento muito importante e sempre presente em MS é “quando os estudos foram publicados?”. Ao extrair essas informações, torna-se possível analisar o crescimento do interesse nesses tópicos de pesquisa ao longo do tempo. Tais informações, quando combinadas com outras, têm o potencial de retratar se a maturidade dos estudos envolvendo esses tópicos tem aumentado ao longo dos anos, além de identificar lacunas e revelar tendências futuras nessas áreas. Por todos esses motivos, uma sexta pergunta foi acrescentada às questões de pesquisa. A **Tabela 1** mostra a correlação entre cada questão de pesquisa e os aspectos que fundamentaram sua elaboração.

Tabela 1. Questões de pesquisa

#	Questão	Referencial considerado
Q1.	Qual a evolução histórica das publicações sobre EDM e LA no Brasil?	Felizardo et al. (2017)
Q2.	Quais as principais contribuições das publicações brasileiras sobre IE para as áreas de EDM e LA?	Tipo de contribuição (PETERSEN et al., 2015)
Q3.	Que tipos de pesquisas envolvendo EDM e LA têm sido feitas no Brasil?	Tipo de pesquisa (PETERSEN et al. 2015; WIERINGA et al., 2005)
Q4.	Em quais ambientes os artigos brasileiros sobre EDM e LA têm sido desenvolvidos?	Foco de estudo (PETERSEN et al. 2015)
Q5.	Quais os métodos de pesquisa adotados nesse tipo de publicação?	Método de pesquisa (PETERSEN et al., 2015; WOHLIN et al., 2012; EASTERBROOK et al., 2008)
Q6.	Quais os canais de divulgação científica onde os estudos têm sido originalmente publicados?	Veículo de publicação (PETERSEN et al. 2015)

A seguir, no quadro da **Tabela 2** são apresentadas as principais motivações para a elaboração das questões de pesquisa:

Tabela 2. Principais motivações

Q1	Tem o propósito de investigar como se deu a progressão histórica das pesquisas voltadas para LA e EDM, publicadas pela comunidade brasileira de IE.
Q2	Trata do tipo de intervenção apresentado que resultou em conhecimento. Essa produção de conhecimento pode ter sido gerada, por exemplo, a partir de uma análise, processo, produto, entre outros.
Q3	Refere-se ao tipo de estudo apresentado. A categorização dos artigos segundo esse critério baseou-se na identificação dos objetivos dos trabalhos, sendo adotada a classificação proposta por Wieringa et al. (2005).

Q4	Está relacionada ao contexto no qual o trabalho foi realizado, podendo ser, de acordo com Felizardo et al. (2017), no âmbito acadêmico, industrial, governamental, de projeto ou organizacional.
Q5	É relativa aos métodos científicos usados. Foi usada a categorização proposta por Easterbrook et al. (2007), na qual as opções são: <i>survey</i> , estudo de caso, experimento controlado, pesquisa ação, etnografia, simulação, prototipagem e análise e prova matemática.
Q6	Corresponde aos canais de divulgação científica onde os estudos têm sido originalmente publicados. A classificação utilizada considera as três categorias básicas citadas por Felizardo et al. (2017), a saber: periódicos (revistas científicas especializadas ou <i>journals</i>), conferências (incluindo os simpósios), e <i>workshops</i> (eventos menores realizados no mesmo espaço físico ou como eventos satélites de conferências)

3.2 ESTRATÉGIA DE PESQUISA

Para as questões definidas na seção anterior, foi realizada uma pesquisa do tipo exploratória por meio de um MS com a análise de artigos científicos em repositórios e sites especializados no tema. O procedimento adotado será detalhado nas seções seguintes.

3.2.1 Repositórios pesquisados

Para a condução deste trabalho, foram feitas buscas automáticas por trabalhos relacionados ao tema de interesse, utilizando como ferramenta de busca as bases de dados de periódicos e eventos nacionais sobre IE disponíveis no portal de publicações da Comissão Especial de Informática na Educação (CEIE). Foram consultadas as bases da Revista Brasileira de Informática na Educação (RBIE)⁹, do Simpósio Brasileiro de Informática na Educação (SBIE)¹⁰, do Workshop de Informática na Escola (WIE)¹¹ e dos Anais dos Workshops do Congresso Brasileiro de Informática na Educação (WCBIE)¹². Estes repositórios foram escolhidos em razão de suas relevâncias para a área de educação e computação.

3.2.2 Palavras-chave utilizadas

Para a pesquisa nos repositórios acadêmicos, foram utilizadas palavras-chaves associadas ao contexto deste estudo, a saber: “*learning analytics*” e “*educational data mining*”, bem como seus respectivos termos correspondentes em português, são eles: “analíticas de aprendizagem” e “mineração de dados educacionais”. Utilizou-se “” (aspas) para delimitar os resultados das pesquisas à ocorrência exata das palavras incluídas entre aspas, buscando criar um filtro mais específico. Vários testes foram realizados iterativamente para compor a *string* de busca. Dessa forma, as buscas foram

⁹ <https://www.br-ie.org/pub/index.php/rbie>

¹⁰ <http://www.br-ie.org/pub/index.php/sbie/index>

¹¹ <http://www.br-ie.org/pub/index.php/wie/index>

¹² <https://www.br-ie.org/pub/index.php/wcbie/index>

realizadas de acordo com a seguinte expressão: ("educational data mining" OR "learning analytics" OR "estatísticas de aprendizagem" OR "mineração de dados educacionais").

3.2.3 Critérios de seleção

Critérios de Inclusão (CI): Como critérios de inclusão das referências bibliográficas, foram considerados trabalhos que: [CI 1]: Sejam estudos primários; [CI 2]: Versam sobre o cenário atual da utilização de LA; [CI 3]: Possuem seus textos integralmente disponíveis na web; [CI 4]: Escritos nos idiomas português ou inglês.

Critérios de Exclusão (CE): Além de consistirem da negação dos critérios de inclusão, foram adotados os seguintes critérios de exclusão: [CE 1]: Estudos redundantes tanto em relação à autoria quanto ao assunto abordado; [CE 2]: Estudos incompletos.

3.2.4 Seleção dos artefatos

A escolha dos artigos foi realizada da seguinte maneira: primeiramente, foi executada a busca automática nas bases bibliográficas. A execução da busca automática resultou em um total de 37 artigos. Sobre esse montante, foi realizada uma triagem para descobrir quais artigos permaneceriam na amostra do universo pesquisado. Para isso, foram aplicados os critérios de seleção definidos na seção anterior.

A triagem foi realizada em duas etapas, a primeira considerou informações disponibilizadas nos metadados que descreviam os artigos científicos e a segunda consistiu na sua leitura completa. Na primeira etapa, foram lidos os títulos, resumos e palavras-chave dos artigos encontrados, informações consideradas relevantes para a categorização dos trabalhos. Quando essa leitura não era considerada suficiente para identificar de imediato se o artigo era relevante, este era separado para leitura completa (segunda etapa). Havendo a seleção do trabalho, suas informações eram registradas no instrumento de análise e passava-se então para o próximo artigo. A metodologia citada para seleção dos artigos foi realizada pelos autores deste trabalho.

A Tabela 3 mostra o quantitativo de estudos retornados por cada base de dados antes e após a triagem. Do total de 37 trabalhos, 10 foram descartados, totalizando 27 estudos selecionados. A lista final dos trabalhos selecionados pode ser obtida no endereço: <https://tinyurl.com/ycx3fcjm>.

Tabela 3. Estudos retornados por cada base de dados antes e após a triagem.

Base de dados	Quantidade de estudos	
	Antes da triagem	Após a triagem
RBIE	5	3
SBIE	15	13
WCBIE	17	11

WIE	0	0
Total	37	27

3.3 AMEAÇAS À VALIDADE

Primeiramente, é importante ressaltar que os resultados e contribuições deste trabalho estão limitados ao conjunto de estudos primários encontrados no processo de busca e seleção de artigos, existindo, portanto, o risco inerente de não haver uma correspondência fiel entre o corpus da pesquisa e o material efetivamente relevante envolvendo os tópicos pesquisados. Para diminuir essa possibilidade, variações da *string* de busca foram testadas repetidamente, no intuito de aperfeiçoar a expressão usada em definitivo, conforme descrito na subseção 4.2.2. Similarmente, os critérios de seleção foram elaborados de forma a serem específicos o suficiente para garantir a delimitação adequada sem, contudo, comprometer a abrangência da cobertura desejada (subseção 4.2.3).

Também deve ser pontuada a possibilidade de a quantidade de artigos selecionados ser baixa ao ponto de não haver evidências suficientes para viabilizar a generalização dos resultados. Contudo, considerando a eficácia do mecanismo de busca em filtrar os principais resultados, aliada aos procedimentos da triagem realizada, descrita na subseção 4.2.4, avaliamos que o quantitativo definido foi suficiente para o escopo deste estudo.

Outros riscos potenciais estão relacionados à confiabilidade do instrumento de avaliação e à possível realização de análises tendenciosas por parte dos avaliadores. Sobre o instrumento avaliativo, sua elaboração foi embasada na literatura relevante sobre o tema (subseção 3.2.1), para assegurar maior respaldo científico ao mecanismo adotado. Com relação ao processo de extração de dados, este foi executado de forma supervisionada para garantir uma análise mais acurada e imparcial, reduzindo assim o viés tendencioso, conforme descrito adiante, na subseção 4.4.

3.4 ESTRATÉGIA PARA EXTRAÇÃO DOS DADOS

Os artigos foram distribuídos entre os avaliadores e cada um lido por um autor e mais um autor-revisor, de modo que os dados obtidos pudessem ser validados entre os pares. Havendo divergências, o procedimento metodológico era repetido e discutido em conjunto para se estabelecer o consenso.

4 ANÁLISE E DISCUSSÃO DOS RESULTADOS

A análise dos resultados será descrita a partir das questões de pesquisa definidas anteriormente, portanto, a discussão será orientada pela sequência de respostas às respectivas perguntas.

Q1. Qual a evolução histórica das publicações sobre EDM e LA no Brasil?

A partir dos dados coletados, foi possível ordenar cronologicamente os estudos encontrados. Em relação ao crescimento da quantidade de pesquisas publicadas em IE abordando princípios de EDM ou LA, foram encontrados artigos a publicados partir de 2014, indicando que a exploração desses tópicos em pesquisas brasileiras é relativamente recente. Além disso, foi possível observar que, no geral, houve um aumento significativo nos últimos dois anos do período analisado, revelando uma tendência crescente do interesse nesses tópicos. Contudo, em uma análise mais detalhada, nota-se que esse crescimento se deu de forma variada, quando se olha isoladamente para cada um dos veículos. Um exemplo disso é que, em 2017, houve um declínio desse tipo de publicações no WCBIE e WIE. Por outro lado, destaca-se o crescimento de publicações do SBIE em 2017 envolvendo esses mesmos itens de pesquisa. Um resumo dessa questão pode ser visto no gráfico da **Figura 1**.

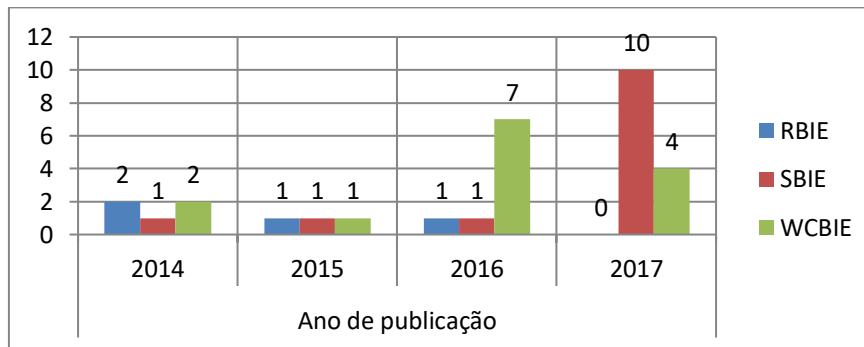


Figura 1. Distribuição temporal das publicações sobre EDM e LA no Brasil

Q2. Quais as principais contribuições das publicações brasileiras sobre IE para as áreas de LA e EDM?

A **Figura 2** mostra os principais tipos de contribuição encontrados. Neste quesito, o esquema de classificação emergiu dos próprios estudos selecionados. É notável, portanto, que as iniciativas são, em sua maioria, focadas em propor ou construir alguma ferramenta, abordagem, método, modelo, arquitetura, algoritmo, técnica, framework ou indicadores educacionais. Em um número levemente inferior, percebe-se que também há iniciativas para os que estão no âmbito de prover conhecimento através de alguma forma de validação, isto é, esses trabalhos não apresentam apenas propostas de solução. Nessa categoria estão os estudos que apresentam soluções sendo usadas na prática na indústria. Entretanto, mesmo em menor proporção, vale ressaltar que muitos trabalhos são voltados para a busca de dados relevantes e convenientes obtidos através de experiências, fornecendo evidências empíricas que permitam, ou não, a sua generalização.



Figura 2. Tipos de contribuição

Q3. Que tipos de pesquisas envolvendo EDM e LA têm sido feitas no Brasil?

O esquema de classificação para tipos de pesquisa adotado foi o proposto por Wieringa et al. (2005) e revisado por Petersen et al. (2015), o qual considera os seguintes tipos de pesquisa: pesquisa de avaliação, pesquisa de validação, proposta de solução, artigo filosófico, relato de experiência e artigo de opinião, conforme pode ser visto no gráfico da Figura 3. Constatou-se que a maioria dos estudos está concentrada em validar a aplicação de técnicas, ferramentas, abordagens e métodos específicos. Propostas de solução e relatos de experiência também foram encontradas, porém em menor quantidade.

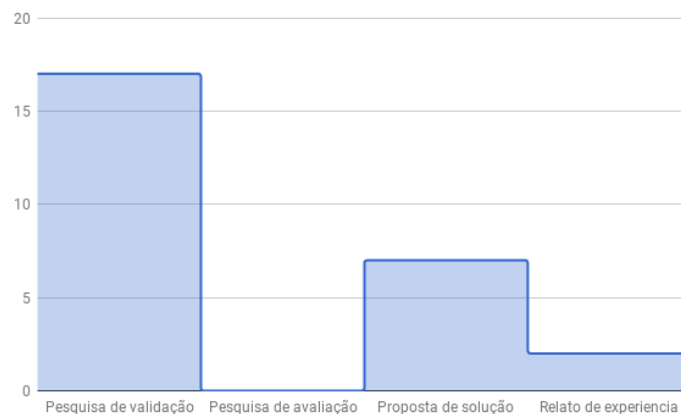


Figura 3. Tipos de pesquisa

Q4. Em quais ambientes os artigos brasileiros sobre EDM e LA têm sido desenvolvidos?

A análise revelou que os estudos têm sido conduzidos predominantemente na academia, isto é, em sua quase totalidade, as pesquisas envolvendo aspectos de EDM e LA são desenvolvidas sem integração com a indústria. Isso evidencia o caráter puramente acadêmico dos pesquisadores brasileiros, o que, por outro lado, dificulta que os veículos de publicação científica do Brasil tenham maior visibilidade e que as pesquisas recebam incentivos financeiros mais expressivos.

Q5. Quais os métodos de pesquisa adotados nesse tipo de publicação?

Foi possível identificar os métodos de pesquisa mais empregados, são eles: estudo de caso e experimento controlado. Verificou-se que 18 trabalhos têm a prática do estudo de caso como procedimento metodológico, esse quantitativo equivale a 64,28% dos estudos primários, isso reflete um objetivo recorrente desses estudos em utilizar EDM e LA para investigar os fenômenos de aprendizagem contemporâneos dentro de seu contexto real. O segundo método mais adotado foi experimento controlado, isto significa que 20% dos trabalhos são focados em delimitar um escopo no qual se possa observar a realidade através de intervenções em determinadas variáveis para identificar os fatores que contribuem para que o tema em questão se materialize. A Figura 4 sintetiza esse resultado.

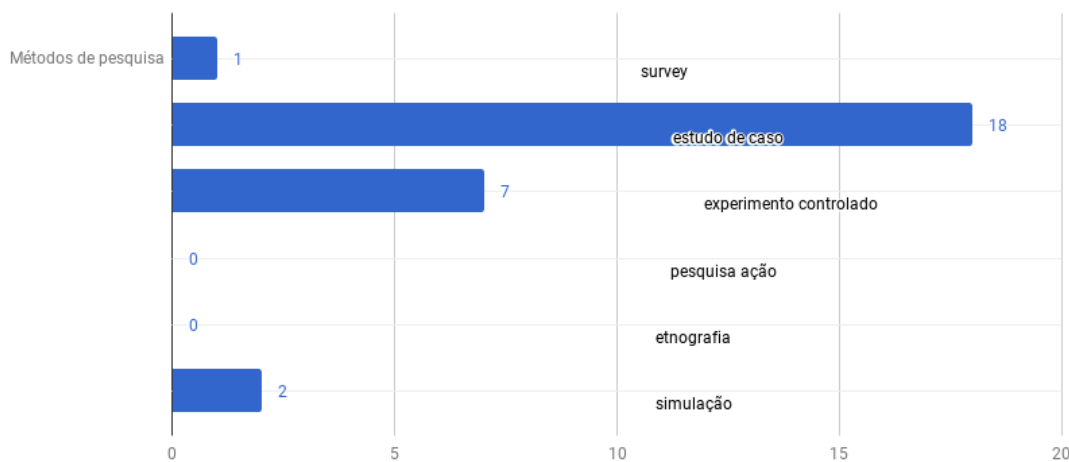


Figura 4. Métodos de pesquisa

Q6. Quais os canais de divulgação científica onde os estudos têm sido originalmente publicados?

Os veículos brasileiros de publicação científica sobre IE que publicaram trabalhos sobre EDM e LA, considerados neste mapeamento, estão listados na **Tabela 3**. Percebe-se, portanto, a ausência de publicações oriundas do WIE e, por outro lado, a grande presença de pesquisas publicadas SBIE e WCBIE, este último sendo destaque, com 17 trabalhos publicados. Também foi possível reparar que a publicação de trabalhos desta natureza tem sido tímida no RBIE.

5 Considerações finais

Esse artigo apresentou um MS das pesquisas envolvendo abordagens de EDM e LA no Brasil. Foram analisados 27 artigos científicos originais publicados em conferências e revistas nacionais relevantes sobre IE. Os estudos foram agregados de acordo com as categorias definidas, no intuito de classificar a literatura relevante para direcionar futuras atividades de pesquisa.

A baixa quantidade de estudos primários encontrados indica que a exploração de técnicas de EDM e LA ainda é recente no Brasil. Contudo, a progressão temporal do volume de publicações atesta que essas são áreas de pesquisa em expansão no cenário nacional. Outra evidência de que os estudos nessas áreas são incipientes foi a descoberta de que a maior parte das iniciativas de pesquisa ainda estão concentradas no âmbito acadêmico, pois maioria dos trabalhos são estudos de caso ou experimentos controlados nos quais são apresentados resultados de experiências realizadas, em sua maioria, em laboratórios de universidades. Há, portanto, a necessidade de explorar novos cenários e contextos, pois acreditamos que há um grande potencial de exploração de outros universos amostrais, dado o volume de dados produzidos em plataformas educacionais no Brasil.

Embora os resultados dessas pesquisas sejam relevantes, percebe-se a necessidade de testificar a efetiva contribuição dos modelos, métodos, técnicas e ferramentas propostas, pois a maioria refere-se a estudos avaliando novas soluções, ainda não implantadas na indústria. Esse fato é compreensível, uma vez que as pesquisas sobre EDM e LA são recentes. Dessa forma, podemos expressar uma visão otimista em relação aos trabalhos futuros.

Por fim, com este mapeamento, acreditamos ter criado um corpo de conhecimento capaz de posicionar novas pesquisas nessas áreas e que a sua utilização como guia de referência pode auxiliar no direcionamento de estudos vindouros, para que esses possam chegar a novas conclusões a partir da maturidade experimental dos estudos aqui mapeados.

REFERÊNCIAS

- BAKER, R. S. J.; INVENTADO, P. S. **Educational Data Mining and Learning Analytics**. In: Larusson, J. A.; White, B.[Eds.]. *Learning Analytics: From Research to Practice*. Springer New York., 2014, p. 61–75.
- BRUSILOVSKKY, P. **Adaptative Hypermedia**. User Modeling and User Adapted Interaction. Kluwer Academic Publishers, 2001.
- EASTERBROOK, S.; SINGER, J.; STOREY, M.; DAMIAN, D. **Selecting Empirical Methods for Software Engineering Research**. Computer Science Department, University of Toronto, 2011.
- FELIZARDO, Katia Romero et al. **Revisão sistemática da literatura em engenharia de software: teoria e prática**. Rio de Janeiro: Elsevier, 2017.
- GOLDSCHMIDT, Ronaldo; PASSOS, Emanuel; BEZERRA, Eduardo. **Data Mining: Conceitos, técnicas, algoritmos, orientações e aplicações**. 2 ed. , Rio de Janeiro, Elsevier, 2015.
- KITCHENHAM, B. A. **Procedures for Performing Systematic Reviews**. Tech. report TR/SE-0401, Keele University, 2004.

MOISSA, B.; GASPARINI, Isabela; KEMCZINSKI, Avanilde. **Learning Analytics**: um mapeamento sistemático. In: TISE2014 - Taller International de Software Educativo - XIX Conferência Internacional sobre Informática na Educação, 2014, Fortaleza. TISE2014 - Taller International de Software Educativo - XIX Conferência Internacional sobre Informática na Educação, 2014. p. 283-290.

_____. **Educational Data Mining versus Learning Analytics**: estamos reinventando a roda? Um mapeamento sistemático. In: XXVI Simpósio Brasileiro de Informática na Educação (SBIE 2015), 2015b, Maceió. Anais... Porto Alegre: Sociedade Brasileira de Computação, 2015, p 1167-1176.

PETERSEN, K.; FELDT, R.; MUJTABA, S.; MATTSSON, M. **Systematic Mapping Studies in Software Engineering**. In *Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering (EASE'08)*. British Computer Society, 2008.

RODRIGUES, R. et al. **A literatura brasileira sobre mineração de dados educacionais**. In: Anais do CBIE. 2014.

VILLARDI, R.; LUCK, E. H.; BARRETO, C. **Introdução à educação à distância**. 1. ed. Rio de Janeiro/Niterói: Sesi/UFF, 2010.

WENGER, E. **Artificial Intelligence and Tutoring Systems**: Computational and Cognitive Approaches to the Communications of Knowledge. Los Altos, CA, Morgan Kaufmann Publishers, 198.

WIERINGA, R.; MAIDEN, N. A. M.; MEAD, N. R.; ROLLAND, C., **Requirements Engineering Paper Classification and Evaluation Criteria**: a Proposal and a Discussion. *Requirements Engineering*, 11(1), 102–107, 2005.