
Mineração de Dados Educacionais e Learning Analytics no contexto educacional brasileiro: um mapeamento sistemático

Educational Data Mining and Learning Analytics in Brazilian educational context: a systematic mapping

Aline de Campos

Universidade Federal do Rio Grande do Sul

Cristiano Galafassi

Universidade Federal do Rio Grande do Sul

Ederson Bastiani

Universidade Federal do Rio Grande do Sul

Fabio Josende Paz

Universidade Federal do Rio Grande do Sul

Raphael Leite Campos

Universidade Federal do Rio Grande do Sul

Leandro Krug Wives

Universidade Federal do Rio Grande do Sul

Sílvio César Cazella

Universidade Federal do Rio Grande do Sul e
Universidade Federal de Ciências da Saúde de Porto Alegre

Eliseo Berni Reategui

Universidade Federal do Rio Grande do Sul

Dante Augusto Couto Barone

Universidade Federal do Rio Grande do Sul

Resumo: A partir de um mapeamento sistemático buscou-se verificar as contribuições de *Learning Analytics* e Mineração de Dados Educacionais no contexto educacional brasileiro. Optou-se por fontes de busca em três revistas na área de Informática na Educação e Anais de dois eventos de relevância nacional onde foram verificados 136 artigos entre janeiro de 2008 e março de 2020. Após a aplicação dos critérios de exclusão e de qualidade foram selecionados 71 artigos. Os resultados apresentam a maior ocorrência de projetos com finalidade de analisar desempenho acadêmico e prevenção de evasão escolar, embora nos últimos anos os assuntos estejam apresentando diversificação temática. Grande parte são voltados ao Ensino Superior e na modalidade de Educação a Distância. Há variedade de tecnologias e recursos utilizados no desenvolvimento de soluções em LA ressaltando o uso de Linguagem R, MySQL e a ferramenta Weka.

Palavras-chave: mapeamento sistemático, mineração de dados educacionais, *learning analytics*.

Abstract: From a systematic mapping, Learning Analytics (LA) and Educational Data Mining (EDM) contributions in the Brazilian educational context were verified. The sources chosen were three journals in the area of Computers in Education and annals of events of national relevance where 136 articles between 2008 and 2020 were verified. After applying exclusion and quality criteria, 71 articles were selected. Results show the highest occurrence of projects with the purpose of analyzing academic performance and prevention of school dropout, although in recent years the subjects are presenting thematic diversification. Most of them are focused on Higher and Distance Education. There are a variety of technologies and resources used in the development of solutions in Learning Analytics highlighting the use of R Language, MySQL and Weka tool.

Keywords: systematic mapping, educational data mining, learning analytics.

1 Introdução

As mudanças nos processos de ensino e aprendizagem são evidentes nas últimas décadas. Muito disso ocorreu pelos novos desafios e oportunidades pedagógicas e computacionais que surgiram, sobretudo por novas metodologias e pela popularização e crescimento da modalidade de Ensino a Distância (EAD) (QUEIROGA et al., 2015). Com essa realidade, percebeu-se também alunos em constante mudança de perfil, sendo necessárias novas e eficazes ferramentas para auxiliar nesse processo (SIEMENS e LONG, 2011). Portanto, ao longo desse tempo as instituições de ensino vêm produzindo e armazenando grande quantidade de dados. Todavia, ainda hoje são carentes no que tange a sistemas que ofereçam, a partir destes dados, informações rápidas, preditivas e específicas para a realidade administrativa e acadêmica, diminuindo oportunidades de intervenções em tempo real (SIEMENS e LONG, 2011; FILATRO, 2020).

Esta conjuntura abriu espaço para que se desenvolvessem ao longo dos anos ações relacionadas ao uso de dados em âmbito educacional. As áreas de *Educational Data Mining* (EDM, em português Mineração de Dados Educacionais) e *Learning Analytics* (LA, que pode ser traduzida como Analítica de Aprendizagem), por exemplo, têm sido aprofundadas com o uso de diferentes conjuntos de técnicas e recursos tecnológicos capazes de prover indicações relevantes aos aspectos pedagógicos, didáticos e gerenciais no âmbito educacional.

Nesse contexto, técnicas de medição, coleta, análise e interpretação dos dados permite, entre outras questões, avaliar o progresso acadêmico, prever e indicar cenários e identificar possíveis problemas, além de auxiliar no acompanhamento e visualização de aspectos fundamentais nos processos, tanto de ensino, quanto de aprendizagem (SIEMENS e LONG, 2011; RIGO et al., 2014; FILATRO, 2020).

Frente a esse cenário, o presente artigo apresenta a concepção, desenvolvimento e resultados de um Mapeamento Sistemático de Literatura (MSL) que teve como objetivo identificar quais ferramentas e recursos baseados em EDM e LA estão sendo aplicados no contexto educacional brasileiro, a fim de oferecer subsídios a novas pesquisas, detectar lacunas e potencialidades de aprofundamento nessas áreas.

Este artigo está estruturado em cinco seções, incluindo a introdução. Na seção 2 apresenta-se uma breve fundamentação sobre Mineração de Dados Educacionais e *Learning Analytics* e os seus desdobramentos em técnicas, recursos e aplicações. Na seção 3, os materiais e métodos utilizados são apresentados. A seguir, na seção 4, os resultados deste MSL são discutidos. Por fim, apresentam-se as considerações finais, na seção 5, e as referências bibliográficas que suportaram este estudo.

2 Correlações e distinções entre Mineração de Dados Educacionais e Learning Analytics

As áreas de EDM e LA estão diretamente relacionadas. Ambas têm como objetivo a melhoria de processos de ensino e aprendizagem por meio da análise de dados em larga escala, de maneira sistematizada, que possam auxiliar na ampliação de processos de avaliação, compreensão de problemas e planejamento de intervenções (SIEMENS e BAKER, 2012; FILATRO, 2020). Porém,

embora relacionadas, segundo Filatro (2020) "EDM e LA, têm orientações tecnológicas, ideológicas e metodológicas distintas" (p. 84).

A Mineração de Dados Educacionais (*Educational Data Mining*) ocupa-se, principalmente, dos processos de coleta e exploração dos diversos tipos de dados encontrados em ambientes educacionais para melhor compreensão do contexto de aprendizagem dos alunos (MOISSA e KEMCZINSKI, 2015). Em se tratando de dinâmica, costuma seguir uma sequência de passos, com a descoberta de informação, utilizando-se de recursos automatizados numa abordagem com ênfase na redução de problemas em componentes menores, a fim de analisá-los tanto individualmente, quanto em suas relações (SIEMENS e BAKER, 2012).

Já a área de analítica de aprendizagem (*Learning Analytics*) busca coletar, medir, analisar e relatar os dados e seus contextos, com objetivo de otimizar o aprendizado e o ambiente em que este ocorre (MOISSA e KEMCZINSKI, 2015). Em seu processo, apresenta ciclos que consideram a intervenção e participação humana. Sendo assim, procura-se entender o sistema como um todo, numa abordagem holística (SIEMENS e BAKER, 2012). Ou seja, LA "está muito mais ancorado na intervenção humana" do que outras abordagens (FILATRO, 2020, p. 83).

Um processo de LA tem início na coleta de dados. Na sequência, ocorre um processo de mineração de dados e seus resultados, geralmente, são apresentados em um *dashboard* integrado ao ambiente de aprendizagem, para que os usuários, geralmente professores, possam obter informações que os auxiliem nas tomadas de decisões acadêmicas (DYCKHOFF et al., 2012). Acrescenta-se que LA pode ser usada para avaliar currículos, programas e instituições, auxiliando a fornecer uma análise mais profunda, bem como para transformar a pedagogia de uma forma mais radical (JOHNSON et al., 2011).

De acordo com Bienkowski et al. (2012) LA permite responder a questões como:

- a) Quando os alunos estão prontos para avançar para o próximo tópico?
- b) Quando é que os estudantes ficam para trás numa determinada unidade curricular?
- c) Quando é que um aluno está em risco de não concluir um curso/unidade curricular?
- d) Que nota um estudante obterá sem intervenção?
- e) Qual é o melhor curso que um determinado estudante pode optar a seguir?
- f) Um estudante deve ser encaminhado a um conselheiro, a fim de obter ajuda?

Há grande demanda por estudos avançados sobre o uso de dados educacionais no sentido de ampliar o desenvolvimento de novos processos e ferramentas voltadas para melhorias na aprendizagem, tanto para alunos, quanto de apoio aos docentes. Além disso, se faz relevante a integração de ferramentas de EDM e LA em processos de ensino e aprendizagem como ferramenta pedagógica, e não apenas como recurso secundário (ELIAS, 2011). Além disso, os resultados gerados por processos de mineração de dados e sistematização de informações podem ser úteis na gestão educacional em diversos níveis, produzindo conhecimento sobre o contexto para auxiliar na tomada de decisões.

Enfim, entende-se que tanto EDM, quanto LA, com suas particularidades e correlações, são capazes de representar "uma forte voz e força para a excelência do Data Science na Educação,

orientando os formuladores de políticas, administradores e educadores na implementação de boas práticas de educação orientada a dados" (FILATRO, 2020, p. 85).

Tendo em vista este cenário, optou-se pela realização de um mapeamento sistemático de literatura, no sentido de investigar a produção brasileira a respeito desta temática, com foco em detectar aplicações de EDM e LA e perceber os direcionamentos adotados, bem como potenciais lacunas para estudo.

3 Materiais e métodos

O foco deste mapeamento sistemático de literatura é identificar as contribuições de Mineração de Dados Educacionais e *Learning Analytics* no contexto educacional brasileiro. Optou-se por este tipo de estudo pelo desejo de obter uma visão geral das duas áreas, sobretudo no que diz respeito às tecnologias e recursos utilizados. Em se tratando de revisões de literatura, o mapeamento sistemático apresenta-se como o processo de coleta, avaliação e sistematização de trabalhos primários, apresentando uma abordagem mais ampla e sobretudo quantitativa, capaz de prover dados para identificação de tendências de estudos, bem como lacunas a serem exploradas (PETERSEN et al., 2008). Assim, optou-se pela realização de um mapeamento sistemático, por tratar-se de assuntos amplos e com produção crescente no cenário brasileiro.

Foram definidas as seguintes etapas, baseadas na metodologia proposta por Kitchenham (2004):

- a) planejamento da pesquisa: criou-se um protocolo de pesquisa que leva em conta as questões norteadoras para condução das coletas, critérios de exclusão e qualidade, bem como elementos de extração de dados para análise, com o objetivo de oferecer reprodutibilidade a esta pesquisa, um dos elementos essenciais de um mapeamento sistemático;
- b) seleção de estudos primários: tendo em vista o protocolo estabelecido, procedeu-se a coleta e seleção de estudos de forma criteriosa e sistematizada;
- c) avaliação da qualidade: além dos critérios de exclusão, os artigos que não atenderam à maioria dos critérios de qualidade foram excluídos do estudo, restando assim o conjunto de dados a serem analisados em maior profundidade;
- d) extração de dados: este processo deu-se com a leitura dos artigos selecionados e a extração de informações, conforme os atributos pré-definidos sistematizados em uma tabela;
- e) análise dos resultados: os dados inseridos na tabela de extração foram consolidados em forma de gráficos e nuvem de termos, a fim de possibilitar a análise segundo as questões norteadoras definidas.

A seguir, apresenta-se o percurso da concepção e execução da pesquisa, tendo em vista a metodologia adotada.

3.1 Planejamento da pesquisa

O planejamento da pesquisa engloba o objetivo e a necessidade da mesma, bem como a elaboração das estratégias de busca e de exclusão de estudos primários.

3.1.1 Questões de pesquisa

A questão de pesquisa principal (QP) que este estudo busca responder pode ser enunciada da seguinte forma: "Quais ferramentas e metodologias de mineração de dados educacionais e *learning analytics* estão contribuindo para o processo de ensino-aprendizagem no contexto educacional brasileiro?". A partir dessa questão, outras três questões pontuais foram identificadas como relevantes, as quais podem ser observadas na Tabela 1.

Tabela 1. Fontes de busca para a pesquisa

Código	Questão
Q1	Quais são os principais tópicos de interesse das pesquisas?
Q2	Para quais níveis de ensino as ferramentas são propostas?
Q3	Em que modalidades de ensino essas ferramentas vêm sendo aplicadas?

Fonte: elaborada pelos autores.

3.1.2 Processo de busca

A distinção entre mapeamentos sistemáticos e outros tipos de revisões bibliográficas está no rigor do processo de busca (KITCHENHAM, 2013). Assim, como parte da metodologia, definiu-se a seguinte *string* de busca: **(Learning analytics) OR (Mineração de dados educacionais)**. Ou seja, optou-se por buscar pelos dois termos, usando o operador condicional "OU", uma vez que LA relaciona-se diretamente com EDM, mas ambos podem ser utilizados de forma isolada.

Alguns autores consideram LA como parte fundamental de EDM (ROMERO e VENTURA, 2013), enquanto outros consideram que LA avança EDM com o uso centrado no julgamento humano (SIEMENS e BAKER, 2011). De modo a encontrar o cenário do contexto do ensino brasileiro, a *string* de busca foi submetida às fontes de busca apresentadas na Tabela 2.

Tabela 2. Fontes de busca para a pesquisa

Sigla	Fonte de busca
IEPT	Revista Informática na Educação: Teoria e Prática ¹
RBIE	Revista Brasileira de Informática na Educação ²
RENOTE	Revista Novas Tecnologias na Educação ³
SBIE	Simpósio Brasileiro de Informática na Educação ⁴
WIE	Workshop de Informática na Escola ⁵
WCBIE	Workshops do Congresso Brasileiro de Informática na Educação ⁶

¹ Disponível em <https://seer.ufrgs.br/InfEducTeoriaPratica>

² Disponível em <https://www.br-ie.org/pub/index.php/rbie/index>

³ Disponível em <https://seer.ufrgs.br/renote>

⁴ Disponível em <https://www.br-ie.org/pub/index.php/sbie/index>

⁵ Disponível em <https://www.br-ie.org/pub/index.php/wie>

⁶ Disponível em <https://www.br-ie.org/pub/index.php/wcbie/index>

JAIE	Jornada de Atualização em Informática na Educação ⁷
DESAFIE	Desafios da Computação Aplicada à Educação ⁸

Fonte: elaborada pelos autores.

Foram usadas como fontes de pesquisa três revistas brasileiras com temática relacionada à Informática na Educação, bem como eventos componentes do Congresso Brasileiro de Informática na Educação (CBIE), que tem grande relevância e abrangência nacional realizado anualmente há mais de 20 anos. Essas fontes foram escolhidas sobretudo por sua qualificação perante a comunidade acadêmica e por apresentarem todo o seu acervo de artigos com processo de revisão por pares, o que amplia o grau de confiabilidade das publicações.

A Revista Informática na Educação: Teoria e Prática e a Revista Novas Tecnologias na Educação possuem ambas Qualis B1 nas áreas de Ensino e Interdisciplinar e a Revista Brasileira de Informática na Educação possui Qualis B1 em Educação. Quanto aos eventos, o Simpósio Brasileiro de Informática na Educação possui classificação B1 no Qualis Conferências na área de Ciência da Computação, enquanto o Workshop de Informática na Escola possui Qualis B3 e representa um espaço interessante de apresentação de iniciativas do uso de tecnologias em âmbito educacional correspondente a Educação Básica. Além disso, também foram usadas como fontes os demais eventos desse Congresso: os Workshops, a Jornada de Atualização em Informática na Educação e os Desafios da Computação Aplicada à Educação.

A Tabela 3 apresenta a quantidade de publicações apresentadas a partir da *string* de busca em cada uma das fontes de pesquisa.

Tabela 3. Quantidade de artigos encontrada em cada fonte

IEPT	RBIE	RENOTE	SBIE	WIE	WCBIE	JAIE	DESAFIE
6	11	22	50	2	41	2	2

Fonte: elaborada pelos autores.

Tendo em vista a seleção destas fontes, procedeu-se a busca usando a mesma *string* em todas as ocasiões onde ao todo foram encontrados 136 trabalhos, os quais foram submetidos aos critérios de exclusão.

3.1.3 Critérios de exclusão

Os critérios de exclusão possuem como objetivo remover os trabalhos que não se enquadrem no escopo deste mapeamento. Estes critérios foram elaborados tendo em vista um recorte que pudesse oferecer uma seleção de conteúdos relevantes para colaborar com esta pesquisa. Desta forma, foram definidos 5 critérios de exclusão, apresentados na Tabela 4.

Tabela 4. Critérios de exclusão

Código	Critério
--------	----------

⁷ Disponível em <https://www.br-ie.org/pub/index.php/pie>

⁸ Disponível em <https://www.br-ie.org/pub/index.php/desafie>

CE1	Trabalhos publicados antes do ano de 2008.
CE2	Trabalhos que apresentam revisões ou mapeamentos sistemáticos.
CE3	Trabalhos que não apresentam as ferramentas utilizadas, resultados alcançados ou a metodologia aplicada.
CE4	Trabalhos que não apresentam as palavras-chave no título ou resumo/abstract.
CE5	Trabalhos que não foram aplicados ao contexto educacional brasileiro.

Fonte: elaborada pelos autores.

Destaca-se que o recorte de tempo estabelecido foi do escopo de janeiro de 2008 até março de 2020. Após a aplicação dos critérios de exclusão nos 136 trabalhos, obteve-se o total de 73 trabalhos selecionados (Tabela 5), os quais foram selecionados para leitura na íntegra, de forma a permitir a avaliação da qualidade. Ressalta-se que alguns trabalhos atenderam mais de um critério de exclusão.

Tabela 5. Artigos excluídos pelos critérios de exclusão

	C1	C2	C3	C4	C5	Final
Excluídos	2	13	24	19	5	63 excluídos
Total parcial	134	121	97	78	73	73 selecionados

Fonte: elaborada pelos autores.

3.1.4 Avaliação de qualidade

Para que um trabalho fosse considerado relevante foram estabelecidos critérios de qualidade, para os quais a obra deveria responder de forma positiva a, pelo menos, três dos seguintes questionamentos que constam na Tabela 6.

Tabela 6. Critérios de avaliação da qualidade

Código	Critério
AQ1	O trabalho aborda com clareza a justificativa do estudo?
AQ2	O trabalho aborda com clareza a metodologia?
AQ3	O trabalho aborda com clareza as ferramentas utilizadas?
AQ4	O trabalho aborda com clareza o público alvo (nível e modalidade de ensino)?
AQ5	O trabalho aborda com clareza os resultados alcançados?

Fonte: elaborada pelos autores.

A partir da análise tendo em vista os critérios de qualidade, 2 trabalhos foram removidos da revisão, pois não atenderam à metodologia estabelecida. Logo, 71 trabalhos foram identificados como relevantes para este estudo.

3.1.5 Procedimentos de extração de dados

Após a seleção dos artigos, levando em consideração os critérios de exclusão e de qualidade, criou-se uma tabela de avaliação e extração de dados relevantes capazes de auxiliar nas respostas às questões norteadoras do projeto.

Essa dinâmica foi adotada no sentido de sistematizar os procedimentos de análise de uma forma integrada e convergente, uma vez que este processo foi realizado por múltiplos avaliadores. Após o mapeamento, houve processo de revisão da extração dos dados dos artigos no sentido de oferecer maior assertividade ao processo de coleta. Os elementos verificados foram definidos previamente e são apresentados na Tabela 7.

Tabela 7. Elementos extraídos dos artigos

Elementos	Descrição
Plataforma	Tipo de plataforma onde o sistema é disponibilizado ou aplicado, podendo ser uma integração com o ambiente Moodle, um sistema próprio ou outro sistema educacional.
Tecnologia	Tecnologias computacionais utilizadas no desenvolvimento da estrutura de LA.
Recursos	Indicação dos recursos integrados à estrutura tais como algoritmos, APIs, frameworks.
Tipos de dados	Refere-se à fonte dos dados utilizados na estrutura da solução.
Finalidade	Apresenta a finalidade da solução apresentada no artigo.
Modalidade	Indica a modalidade de ensino para a qual a solução estava indicada no artigo (Presencial, Semipresencial e Educação a Distância).
Nível de Ensino	Indica o nível de ensino para o qual a solução estava indicada no artigo (Ensino Fundamental, Médio, Superior e Pós-graduação).
Público Alvo	Apresenta o público alvo da solução (Alunos, Professores e Gestores).

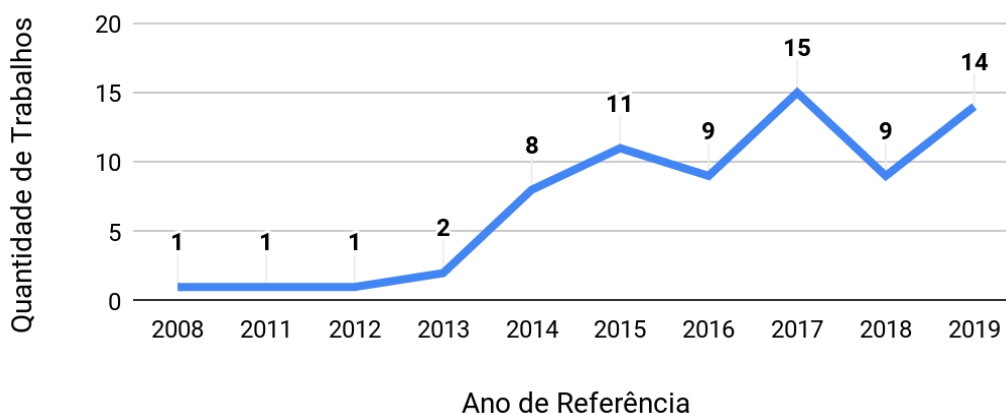
Fonte: elaborada pelos autores.

4 Análise de resultados

Nesta seção serão abordados os resultados da extração e síntese de informações e a análise de dados deste mapeamento sistemático, de forma a responder às questões de pesquisas.

A Figura 1 apresenta a quantidade de trabalhos selecionados por ano de publicação. Percebe-se uma ascensão na quantidade de publicações na temática nos últimos anos.

Figura 1. Quantidade de trabalhos por ano de publicação



Fonte: elaborada pelos autores.

De fato, o uso de dados para apoiar a área educacional tem tido maior produção e visibilidade na última década, em âmbito global. Neste mapeamento percebe-se um crescimento nos últimos cinco anos. Ainda existe um foco grande de desenvolver ambientes virtuais voltados à aprendizagem no sentido de produzir dados. Há uma riqueza de informações existentes nestes ambientes produzidos ao longo das últimas décadas que precisa ser mais bem explorada.

4.1. Q1: Quais são os principais tópicos de interesse das pesquisas?

Para responder a esta questão de pesquisa, este mapeamento sistemático procurou identificar quais eram os tópicos de interesse de cada trabalho. Os artigos foram agrupados de acordo com 5 tópicos amplos, e um tópico genérico (o qual engloba tópicos com menos de três artigos), a partir dos objetivos que os autores apresentaram nos trabalhos. É importante salientar que alguns trabalhos abordam mais que um tópico (Figura 2).

Figura 2. Tópicos de interesse dos trabalhos.



Fonte: elaborada pelos autores.

O principal tópico de interesse dos pesquisadores está relacionado à medição ou melhoria do desempenho acadêmico dos alunos, o qual foi abordado em 36,62% (26/71) dos artigos selecionados. Esta é de fato uma preocupação predominante na área educacional e existem muitas iniciativas buscando apoiar esta questão.

Em seguida, o problema da evasão escolar, seja para análise de motivos ou para prevenção, foi foco de estudo em 21,13% dos trabalhos, isto é, em 15 dos 71 trabalhos. Após, 8,45% dos estudos buscaram aplicar técnicas para classificação de perfis de alunos de acordo com suas participações em ambientes de ensino e aprendizagem. A pesquisa relacionada a aspectos de motivação e engajamento de alunos, por sua vez, representa 7,04% dos estudos.

As interações de tutores em ambientes virtuais de aprendizagem foi o quinto tópico mais pesquisado, representando 4,23% dos artigos classificados. Por fim, os 22,54% (16/71) restantes dos trabalhos foram agrupados para a interpretação por representarem uma pequena parcela de trabalhos com assuntos diversos. Tais trabalhos focaram nos seguintes assuntos apresentados na Tabela 8.

Tabela 8. Assuntos dos trabalhos classificados no grupo "Outros"

Assunto	Número
Visualização de dados	2
Adaptação de ambientes de aprendizagem	2
Interação social	1
Recomendação de ações ou recursos pedagógicos	2
Integração de dados educacionais	1
Docentes com curso superior no ensino fundamental	1
Predição de atividades	1
Melhoria de interação com AVA	3
Comparação de técnicas de análise de dados	1
Verificação de habilidades de leitura e escrita	1
Gestão educacional	1

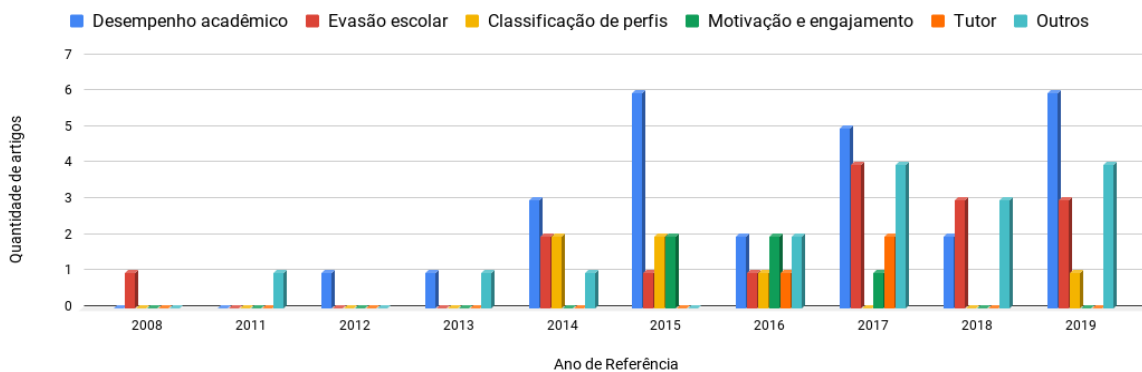
Fonte: elaborada pelos autores.

Percebe-se alguns elementos interessantes, porém ainda com pouca exploração no cenário brasileiro, tais como sistemas de recomendação, predição e verificação de habilidades. Estes parecem ser bons elementos a serem aprofundados em novas pesquisas.

A Figura 3 demonstra a distribuição temática em relação aos anos de publicação. Pode-se verificar a grande ocorrência de enfoque em desempenho escolar de 2012 a 2015 e, embora

ainda com predominância deste tema, uma maior diversificação temática a partir de 2016, o que pode indicar a ampliação de escopo de projetos voltados às áreas pesquisadas. Além disso, verifica-se que a evasão escolar tem sido tópico constante de estudos.

Figura 3. Temática dos trabalhos por ano de publicação.

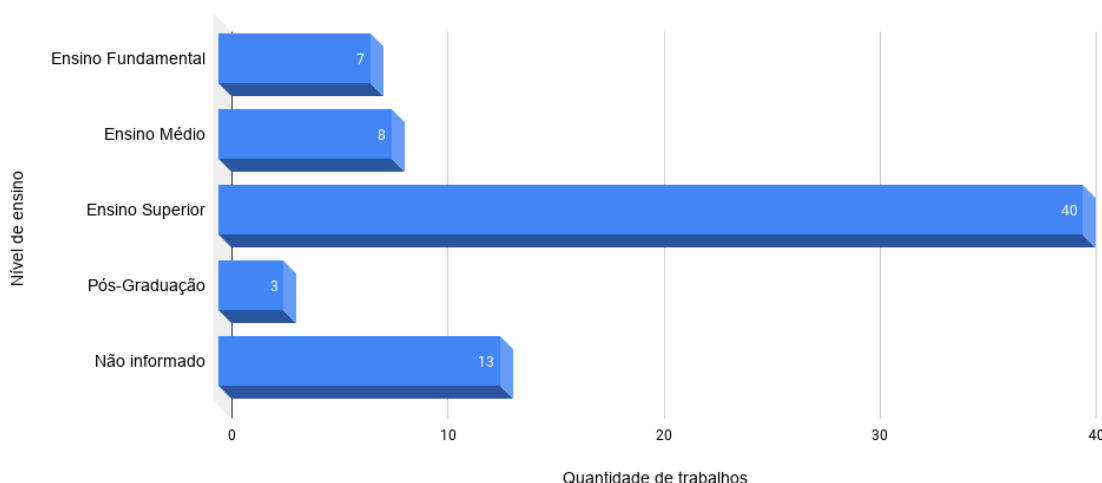


Fonte: elaborada pelos autores.

4.2. Q2: Para quais níveis de ensino as ferramentas são propostas?

As respostas para esta questão procuraram identificar quais são os níveis educacionais atendidos pelas ferramentas de Mineração de Dados Educacionais e *Learning Analytics*. A Figura 4 apresenta o resultado quantitativo para esta questão.

Figura 4. Quantidade de trabalhos por nível de ensino.



Fonte: elaborada pelos autores.

A maioria dos autores foca seus esforços de pesquisa no ensino superior, correspondendo a 56,34% dos trabalhos. Uma hipótese para esta predominância pode estar associada ao fato de

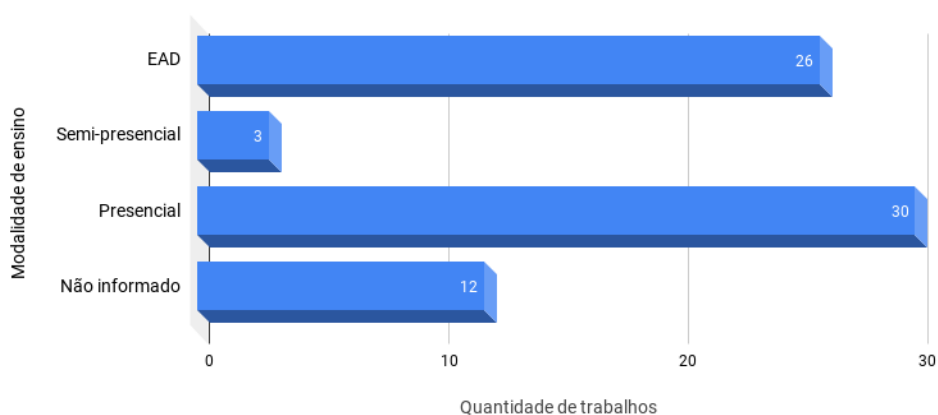
grande parte da produção acadêmica nacional ser produzida no âmbito da Pós-graduação e por facilidade de acesso concentrar-se no Ensino Superior.

Em relação à educação básica, que inclui o Ensino Fundamental (séries iniciais e séries finais) e o Ensino Médio, observa-se 7 trabalhos relacionados ao Ensino Fundamental e 8 estudos relacionados ao Ensino Médio, correspondendo a 21,13% do total analisado. Já o nível de pós-graduação corresponde a apenas 4,23% dos estudos. Por fim, é necessário ressaltar que 13 trabalhos, 18,31%, não informaram para qual nível de estudo suas pesquisas são direcionadas, entretanto apresentavam boa qualidade e uma abrangência de aplicação em diversos níveis de ensino.

4.3. Q3: Em que modalidades de ensino essas ferramentas vêm sendo aplicadas?

Por fim, a terceira questão de pesquisa observa em quais modalidades de ensino estão concentrados os estudos sobre o tema. É possível verificar, através da Figura 5, que há um equilíbrio entre as modalidades de Educação a Distância e Ensino Presencial.

Figura 5. Quantidade de trabalhos por modalidade de ensino.



Fonte: elaborada pelos autores.

A maioria dos autores direciona suas pesquisas para a modalidade de Educação a Distância, compreendendo 36,62% dos trabalhos analisados. Esse resultado já poderia ser esperado, uma vez que essa modalidade exige o intermédio de um ambiente virtual, que por sua vez proporciona uma grande produção de dados que abrem espaço para análise.

O Ensino Presencial está associado a 42,25%. Ainda, 4,23% dos trabalhos são focados no ensino semipresencial. Os outros 16,90% dos autores não informaram a qual modalidade de ensino seus trabalhos são direcionados.

4.4. Análises gerais dos dados coletados

Durante o processo de extração dos dados para realização do mapeamento, como mencionado anteriormente, também foram analisados os critérios: público alvo, tipos de dados utilizados, bem como tecnologias e recursos para o desenvolvimento das soluções. A fim de apresentar

neste artigo uma melhor sistematização e visualização destes dados, optou-se pela criação de nuvens de palavras capazes de demonstrar os termos verificados, bem como suas frequências.

Nesse sentido, fez-se uso de uma ferramenta para construção deste tipo de recurso visual⁹, a qual gera nuvens de palavras a partir da inserção de dados e da frequência de ocorrência dos termos. Para isso, os termos foram extraídos pelos autores deste trabalho, conforme descrito na seção 3.1.5 e, posteriormente, inseridos na ferramenta para geração da visualização. Assim, a seguir apresenta-se o mapeamento para estes critérios elencados anteriormente com destaque para os termos com maior frequência que aparecem em tamanho maior, enquanto os demais termos com menor ocorrência são apresentados em tamanho menor.

Em relação a outros elementos analisados no mapeamento sistemático percebeu-se que o público alvo dos projetos varia no foco em tutores, gestores, docentes e alunos, havendo maior predominância nesses últimos dois públicos. A Figura 6 demonstra nuvens de palavras indicando a frequência de termos.

Figura 6. Visão geral sobre público alvo



Fonte: elaborada pelos autores.

Já em relação ao tipo de dados (Figura 7) usados nos processos, encontra-se uma diversidade que corresponde a dados de alunos com maior ocorrência, mas também dados de docentes e tutores. Dados mais gerenciais como escolares, de turmas e educacionais também foram mencionados.

⁹ Ferramenta para construção de nuvem de palavras *Word Art Cloud Creator* disponível em: <https://wordart.com>

Figura 7. Visão geral sobre tipos de dados utilizados



Fonte: elaborada pelos autores.

Ressalta-se o indício de dados socioeconômicos sendo utilizados em alguns casos, o que pode ser rico no sentido de estabelecer correlações e impactos entre os processos de aprendizagem e as questões sociais e econômicas do estudante e/ou âmbito acadêmico.

Um dos intuítos da realização deste mapeamento sistemático também residiu na possibilidade de ampliar conhecimentos em relação às tecnologias e recursos computacionais que estão sendo utilizados em soluções e Mineração de Dados Educacionais e *Learning Analytics*. Sendo assim, avaliou-se nos artigos selecionados a menção a respeito destes elementos.

Quanto às tecnologias percebe-se grande frequência de projetos que fazem uso da Linguagem de Programação R, da Linguagem de Programação PHP e do sistema gerenciador de banco de dados MySQL. A Linguagem R tem sido bastante explorada no âmbito da análise de dados por conta de sua versatilidade; enquanto a linguagem PHP, em conjunto com o uso de banco de dados relacional MySQL, apresenta grande relevância por sua facilidade de integração com grande parte dos ambientes virtuais de aprendizagem mais populares, além de ampla documentação e comunidade de desenvolvedores. A Figura 8 apresenta em nuvem de palavras as ocorrências para tecnologias verificadas nos artigos selecionados.

Figura 8. Visão geral sobre tecnologias



Fonte: elaborada pelos autores.

Algumas outras tecnologias, como Python e Javascript, apresentaram boa ocorrência. Estas são linguagens bastante versáteis no que diz respeito a análise de dados. Existem dezenas de bibliotecas de códigos de acesso livre que podem ser utilizadas no desenvolvimento de soluções.

Por fim, no que diz respeito aos recursos computacionais, aqui entendidos como *softwares*, estruturas algorítmicas, *frameworks* e API (*Application Interface Programming*), há bastante diversidade de elementos. A Figura 9 demonstra em uma nuvem de palavras os recursos computacionais verificados neste mapeamento.

Figura 9. Visão geral sobre recursos



Fonte: elaborada pelos autores.

A maior ocorrência residiu no software Weka, um software amplamente conhecido e que possui diversos algoritmos para a realização de tarefas de classificação, associação e clusterização, os quais auxiliam diretamente na construção de soluções de mineração de dados e visualização.

Aparecem com destaque também os softwares Gephi, Ucinet e Pajek para geração de grafos e Análise de Redes Sociais (*Social Network Analysis*). Além disso, percebe-se a ocorrência de

ferramentas para análise de dados, tanto proprietárias, como Microsoft Excel e SPSS, quanto de licença aberta como Rapidminer e Scikit-Learn. Alguns elementos relacionados a tarefas e algoritmos como Apriori, Redes Bayesianas e Árvores de Decisão também são mencionados nos artigos analisados.

5 Considerações finais

Neste trabalho foram apresentados os resultados de um mapeamento sistemático sobre Mineração de Dados Educacionais e *Learning Analytics* e suas contribuições no contexto educacional brasileiro. Foi possível verificar que o desempenho acadêmico e a evasão escolar são os principais tópicos de interesse nas pesquisas e que geralmente são realizadas no Ensino Superior nas modalidades EAD e presencial, com público alvo basicamente de professores e alunos.

Percebe-se a necessidade da ampliação da aplicação de soluções de EDM e LA no âmbito do Ensino Fundamental e Médio, uma vez que a base escolar causa forte impacto no processo de desenvolvimento acadêmico no Ensino Superior. Acredita-se que há necessidade de maior atenção à Educação Básica como forma de proporcionar uma visão mais aprofundada sobre aspectos de ensino e aprendizagem capazes de estimular ações de pensamento crítico e adoção de novas metodologias mais alinhadas com o perfil de alunos da era digital.

Dentre os trabalhos avaliados houve uma dificuldade de identificar tecnologias, recursos e plataforma de acesso dos dados, porém naqueles artigos que informaram destaca-se: a ambiente virtual de aprendizagem o Moodle e o software Weka como ferramenta de tratamento dos dados. Em relação às tecnologias foram encontradas diversas, tais como: Linguagens R, PHP, Java e sistemas gerenciadores de banco de dados SQL, sobretudo MySQL e PostgreSQL.

Como limitações deste mapeamento pode-se considerar a restrição a publicações em âmbito nacional em determinados periódicos e anais de evento, o que se configura como um recorte do estado da arte brasileiro relativo a esta temática. Para trabalhos futuros acredita-se que há possibilidade de selecionar artigos deste mapeamento em que seja possível realizar revisões sistemáticas em um processo qualitativo de análise, com foco na resposta a perguntas mais específicas.

Referências

BIENKOWSKI, M., FENG, M., MEANS, B. Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief. Washington, D.C.: Office of Educational Technology, U.S. Department of Education Department of Education. EUA. 2012.

DYCKHOFF, A. L., ZIELKE, D., BÜLTMANN, M., CHATTI, M. A., Schroeder, U. Design and Implementation of a Learning Analytics Toolkit for Teachers. *Educational Technology & Society*, 15 (3), 58–76, 2012.

ELIAS, T. Learning analytics: Definitions, processes and potential (Report). <http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf>. 2011.

FILATRO, Andrea. Data Science na Educação: presencial, a distância e corporativa. São Paulo: Saraiva, 2020.

JOHNSON, L., R. SMITH, H. WILLIS, A. LEVINE, HAYWOOD, K. Learning Analytics. The 2011 Horizon Report. Austin, Texas: The New Media Consortium. <http://net.educause.edu/ir/library/pdf/HR2011.pdf>. 2011.

KITCHENHAM, B. Procedures for performing systematic reviews. Keele, UK, Keele University, 33(2004):1-26, 2004.

MOISSA, B.; GASPARINI, I.; KEMCZINSKI, A. Educational Data Mining versus Learning Analytics: estamos reinventando a roda? Um mapeamento sistemático. In: Anais do XXVI Simpósio Brasileiro de Informática na Educação, 2015.

QUEIROGA, E. M.; CECHINEL, C; ARAUJO, R. M. Um Estudo do Uso de Contagem de Interações Semanais para Predição Precoce de Evasão em Educação a Distância. Anais dos Workshops do IV Congresso Brasileiro de Informática na Educação (CBIE 2015), p. 1074-1083, 2015.

RIGO, S. J.; CAMBRUZZI, W.; BARBOSA, J. L. V.; CAZELLA, S. C. Aplicações de Mineração de Dados Educacionais e Learning Analytics com foco na evasão escolar: oportunidades e desafios. Revista Brasileira de Informática na Educação, Volume 22, Número 1, p. 132-146, 2014.

ROMERO C., VENTURA, S. Data mining in education. WIREs Data Mining Knowl Discov 2013, 3: 12-27, 2013.

SIEMENS, G.; BAKER, R. S. Learning Analytics and Educational Data Mining: Towards Communication and Collaboration. In: Proceedings of the 2nd international conference on learning analytics and knowledge. ACM. Pages 252-254, 2012.

SIEMENS, G., LONG, P. Penetrating the Fog: Analytics in Learning and Education. EDUCAUSE review, 46 (5), 2011.

Recebido em maio de 2020.

Aprovado para publicação em dezembro de 2020.

Aline de Campos

Doutoranda no Programa de Pós-graduação em Informática na Educação da Universidade Federal do Rio Grande do Sul (UFRGS). E-mail: alinedecampos@gmail.com

Cristiano Galafassi

Doutorando no Programa de Pós-graduação em Informática na Educação da Universidade Federal do Rio Grande do Sul (UFRGS). Professor Assistente na Universidade Federal do Pampa - Campus Itaqui. E-mail: cristianogalafassi@gmail.com

Ederson Bastiani

Doutorando no Programa de Pós-graduação em Informática na Educação da Universidade Federal do Rio Grande do Sul (UFRGS). Professor no Instituto Federal Farroupilha - Campus Panambi. E-mail: edersonbastiani@gmail.com

Fabio Josende Paz

Doutorando no Programa de Pós-graduação em Informática na Educação da Universidade Federal do Rio Grande do Sul (UFRGS). Vice-reitor e Professor na Universidade da Região da Campanha (URCAMP). E-mail: fabiojpaz@gmail.com

Raphael Leite Campos

Doutorando no Programa de Pós-graduação em Informática na Educação da Universidade Federal do Rio Grande do Sul (UFRGS). Professor Auxiliar na Universidade do Vale do Rio dos Sinos (UNISINOS). E-mail: prof.raphaelc@gmail.com

Leandro Krug Wives

Professor Associado no Instituto de Informática e dos cursos de Pós-Graduação em Computação (PPGC) e em Informática na Educação (PPGIE) na Universidade Federal do Rio Grande do Sul (UFRGS). Diretor do Centro Interdisciplinar de Novas Tecnologias na Educação (UFRGS). E-mail: wives@inf.ufrgs.br

Sílvio César Cazella

Professor efetivo nos Programas de Pós-Graduação na Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA) e colaborador do Programa de Pós-Graduação em Informática na Educação da Universidade Federal do Rio Grande do Sul (UFRGS). E-mail: silvioc@ufcspa.edu.br

Eliseo Berni Reategui

Professor da Faculdade de Educação e no Programa de Pós-Graduação em Informática na Educação (PPGIE) da Universidade Federal do Rio Grande do Sul (UFRGS). E-mail: eliseoreategui@gmail.com

Dante Augusto Couto Barone

Professor titular do Instituto de Informática e dos cursos de Pós-Graduação em Computação (PPGC) e em Informática na Educação (PPGIE) na Universidade Federal do Rio Grande do Sul (UFRGS). E-mail: barone@inf.ufrgs.br