

Análise sobre a evolução da mineração de dados educacionais segundo a perspectiva dos principais estudos da área: uma revisão narrativa

I Vanessa Faria de **Souza**
IFRS

I Mayara Faria de **Souza**
UTFPR

RESUMO

Com o aumento da disponibilidade de dados, sobretudo no contexto educacional, surgiram áreas específicas para extração de informações relevantes, como a Mineração de Dados Educacionais (MDE), que integra inúmeras técnicas que dão suporte a captação, processamento e análises desses conjuntos de registros. Essa área tem evoluído constantemente nos últimos 20 anos desde os primeiros trabalhos publicados, e tem conquistado cada vez mais espaço, tanto entre pesquisadores da área da educação, como da ciência de dados. Nesse sentido, esse estudo de revisão narrativa aborda e apresenta os principais aspectos da MDE, seus objetivos e processo, e apresenta o contexto de evolução da área com base em estudos de revisão sistemática de literatura que sintetizaram boa parte dos estudos de MDE das últimas duas décadas.

Palavras-chave: Mineração de Dados Educacionais, Revisão Narrativa, Evolução.

■ INTRODUÇÃO

As organizações de diversos seguimentos têm se mostrado muito eficientes em capturar, organizar e armazenar bases dados de grande dimensão, conseguidos por meio operações diárias ou pesquisas, entretanto a chave para avançar é saber como utilizar adequadamente essa grande quantidade de dados para transformá-la em conhecimentos que possam ser utilizados em suas atividades, sejam elas comerciais ou científicas. Nesse sentido, o conceito de Mineração de Dados (*Data Mining*) está se tornando cada vez mais popular como um método para exploração de informações, que podem apontar conhecimentos novos, que possam orientar em decisões em situações de certeza restrita.

Para Aggarwal (2015) a Mineração de Dados (MD) é definida como:

A mineração de dados é o estudo de coleta, limpeza, processamento, análise e obtenção de informações e ideias úteis de dados. Existe uma grande variação em termos de domínios problemáticos, aplicativos, formulações e representações de dados encontradas em aplicativos reais. Portanto, “Mineração de dados” é um termo abrangente usado para descrever esses diferentes aspectos de processamento de dados (AGGARWAL, 2015).

Com a crescente adoção de MD, essa passou a ser empregada com sucesso também no contexto educacional, auxiliando em diversos cenários, e ficou conhecida como Mineração de Dados Educacionais (MDE). Nesse sentido este estudo tem como objetivo sistematizar estudos que contribuíram para a consolidação da MDE como uma área de pesquisa em potencial expansão, para isso são apresentados primeiramente sua definição, seus objetivos e processo de funcionamento e em seguida é detalhado o contexto de evolução da área.

■ DESENVOLVIMENTO

Nessa seção de desenvolvimento serão abordados os principais aspectos sobre a Mineração de Dados Educacionais, descritos com base nos principais autores da área. Para sintetizar o processo de evolução foram selecionados os estudos de revisão sistemática de literatura mais proeminentes e abrangentes disponíveis entre as bases de dados: IEEE Xplore, ACM Digital Library, ERIC, Scopus e ScienceDirect. Nesse sentido essa seção inicia abordando as definições, objetivos e processo da MDE, em seguido é apresentado o seu contexto de evolução e consolidação.

Definições e Objetivos da Mineração de Dados Educacionais

Nos últimos anos a educação tem se modificado, em decorrência do avanço tecnológico disponível que direcionou a uma instrumentação do setor educacional, tanto em

softwares voltados para o ensino, como na administração digital dos registros acadêmicos pelos gestores das instituições, bem como no uso da internet para a aprendizagem, em especial pela popularização do *e-learning*. Todos esses fatores impulsionaram um crescimento exponencial no volume de dados educacionais, e para se analisar uma grande quantidade de dados, é imprescindível contar com recursos computacionais, caso contrário a tarefa torna-se impraticável.

Dessa forma, as técnicas de mineração de dados estão ganhando cada vez mais importância no setor educacional, pois são uma forma de acompanhar, analisar e avaliar o processo de aprendizagem. Provavelmente, as técnicas de mineração de dados podem fornecer aos formuladores de políticas educacionais modelos para apoiar seus objetivos de aprimorar a eficiência e a qualidade do ensino e da aprendizagem. Além disso, o uso de diferentes técnicas de mineração de dados pode ser visto como base para uma mudança sistêmica, capaz de impactar de maneira positiva nas soluções de problemas específicos das Instituições de Ensino, por exemplo, viabilizando soluções que envolvem a personalização dos ambientes educacionais ou fornecendo suporte para o processo de tomada de decisão no ambiente educacional.

Nesse cenário, destaca-se a MDE que utiliza as técnicas da MD para extrair informações relevantes de conjuntos diversificados de dados educacionais. Segundo a Sociedade Internacional de Mineração de Dados Educacionais¹, a MDE pode ser definida da seguinte forma:

É uma disciplina emergente, preocupada com o desenvolvimento de métodos para explorar dados únicos e cada vez mais em larga escala, provenientes de contextos educacionais e usa esses métodos para entender melhor os alunos e as configurações em que aprendem (EDM, 2020).

Em outras palavras, a MD refere-se a um conjunto de técnicas computacionais para extrair informações de grandes massas de dados, e quando os dados analisados são provenientes de contextos educacionais, chama-se MDE (ROMERO; VENTURA, 2013). Igualmente, De Los Reyes *et al.* (2019) define MDE como uma área voltada ao desenvolvimento de métodos para explorar dados oriundos de ambientes educacionais e utilizá-los para compreender melhor os processos de ensino e aprendizagem. Nessa acepção, Baker, Isotani e Carvalho (2011) alegam que a MDE é definida como a área de pesquisa que tem como foco o desenvolvimento de técnicas para explorar conjuntos de dados coletados em ambientes educacionais. Conforme os autores, a natureza destes dados é mais diversa do que a observada nos dados tradicionalmente utilizados em tarefas de mineração, demandando adaptações e novas abordagens. Ao mesmo tempo, essa diversidade nos dados

1 <http://educationaldatamining.org/>

representa um potencial de implementação de recursos fundamentais para auxílio na melhoria da educação (BAKER; ISOTANI; CARVALHO, 2011; DE LOS REYES *et al.*, 2019; RIGO *et al.*, 2014).

Sendo assim, necessita-se de técnicas e ferramentas que auxiliem na tarefa de verificar, interpretar e relacionar esses dados, com o intuito de gerar conhecimento útil e relevante, o que, segundo De Los Reyes *et al.* (2019) já era um objetivo das técnicas de MD, empregadas para identificar padrões de comportamento e encontrar *insights* que provoquem melhorias em produtos e serviços. Romero e Ventura (2007) elencam questões que diferenciam a MDE da MD em outros domínios:

1. *Objetivos*: que podem se relacionar à pesquisa (a) aplicada, que busca responder questões práticas, por exemplo: como melhorar o processo de aprendizagem; e (b) pura, com a finalidade de por exemplo dar sentido às observações. Na maioria das vezes esses objetivos são difíceis de quantificar e exigem seu próprio conjunto especial de técnicas de medição.
2. *Dados*: em ambientes educacionais, existem muitos tipos diferentes de dados disponíveis para mineração. Esses dados são específicos da área educacional, portanto, possuem informações semânticas intrínsecas, relacionamentos com outros dados, e vários níveis de hierarquia significativa.
3. *Técnicas*: problemas educacionais têm algumas características especiais que exigem que a questão da mineração seja tratada de uma maneira diferente. Embora, a maioria das técnicas tradicionais de MD possam ser aplicadas diretamente, outras não podem e devem ser adaptadas ao problema educacional específico. Exemplo disso, é que em se tratando de cenários comuns de MD, as variáveis são em sua maioria numéricas, tratáveis diretamente por algoritmos de *Machine Learning*, enquanto que em ambientes educacionais a grande maioria é categórica, o que implica esforço em pré-processamento e transformações, para codificar essas variáveis em numéricas, para que então possam ser interpretadas pelos algoritmos.

O processo de Mineração de Dados Educacionais

Além de definir os diferenciais da MDE é importante explicar como é seu funcionamento, pois o processo de MDE não é trivial contendo várias fases. Nesse sentido, será explicado o processo de MD de acordo com Aggarwal (2015), depois será apresentada uma sequência de etapas considerada mais adequada, baseada em pesquisas (BAKER; ISOTANI; CARVALHO, 2011; DE LOS REYES *et al.*, 2019; RIGO *et al.*, 2014; ROMERO; VENTURA, 2013, 2020) específicas de MDE e também em experiências com aplicações.

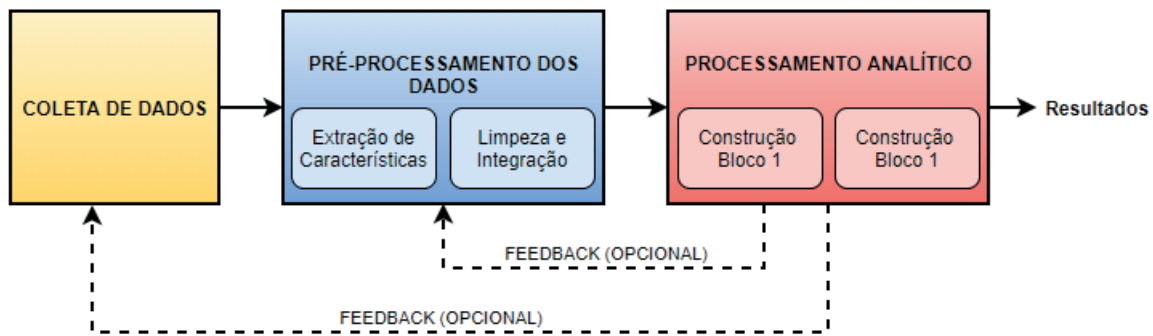
Aggarwal (2015) determina que o fluxo de trabalho de um processo típico de MD contém as seguintes fases:

1. Coleta de dados: A coleta de dados pode exigir o uso de hardware especializado, como uma rede de sensores, trabalho manual, como a coleta de pesquisas com usuários ou ferramentas de software como um mecanismo de rastreamento de documentos da Web para coletar documentos. Esta etapa é específica da plataforma e geralmente fora do domínio do analista de mineração de dados. Após a fase de coleta, os dados geralmente são armazenados em um banco de dados ou, em geral, um *data warehouse*²(tradução livre – armazém de dados) para processamento.
2. Extração de recursos e limpeza de dados (Pré-Processamento e Transformação): Quando os dados são coletados, eles geralmente não estão em um formato adequado para processamento. Para tornar os dados adequados para processamento, é essencial transformá-los em um formato que seja interpretável aos algoritmos de mineração, como multidimensionais, séries temporais ou formato semiestruturado. O formato multidimensional é o mais comum, no qual diferentes campos de dados correspondem às diferentes propriedades medidas que são chamadas de atributos. A fase de extração de recursos geralmente é realizada em paralelo com a limpeza de dados, onde partes ausentes e incorretas dos dados são estimadas ou corrigidas. Em muitos casos, os dados podem ser extraídos de várias fontes e precisam ser integrados em um formato unificado para processamento. O resultado final deste procedimento é um conjunto de dados estruturados, que pode ser efetivamente usado por um programa de computador. Depois da fase de extração de recursos, os dados podem ser armazenados novamente em um banco de dados para processamento.
3. Processamento analítico e algoritmos: A parte final do processo de mineração é projetar métodos analíticos eficazes a partir dos dados processados.

A sequência das etapas do processo de MD proposto por Aggarwal (2015) é apresentada na Figura 01.

² É um repositório central de informações que ficam disponíveis para serem analisadas e dão suporte ao processo de tomada de decisão.

Figura 1. Processo de MD proposto por Aggarwal (2015)



Fonte: Adaptado de Aggarwal (2015)

A partir do processo de MD apresentado, pesquisas em MDE e na experiência com aplicações, algumas das etapas foram ajustadas para condizer efetivamente com o que é executado, levando em conta o fator usuário (programador ou cientista de dados), nesse sentido o processo de MDE pode ser formado pelo seguinte conjunto de etapas:

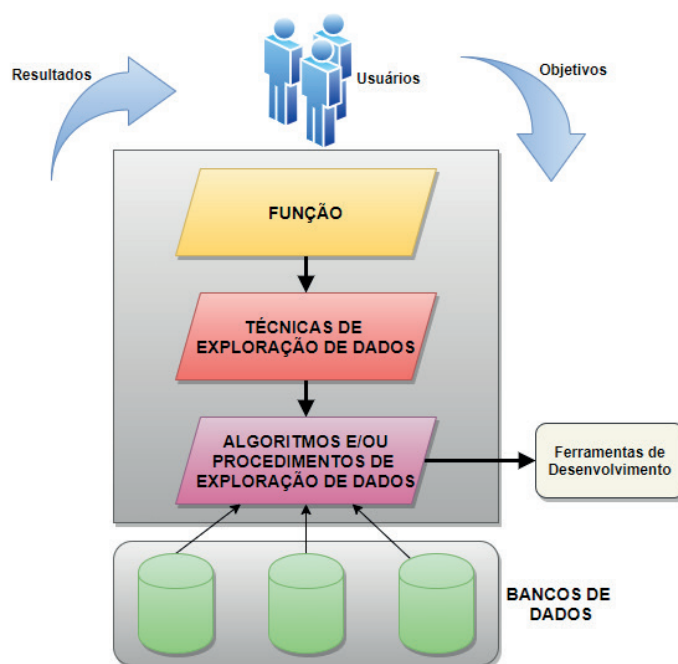
1. Definição da função da MDE: Será realizada a determinação do objetivo do processo MDE, para qual finalidade ela está sendo aplicada, como por exemplo: identificação de padrões, detecção de desvio, segmentação, sistemas de recomendação, análise de ligações e regras de associação, sumarização e visualização, mineração de textos, afinidade em grupos, descrição de grupos, para isso é preciso especificar que tipo de conhecimento pretende-se extrair dos dados.
2. Formatação dos dados que serão utilizados: Diversos tipos de armazenamentos de dados e de bancos de dados podem ser manipulados no processo de mineração, cabe ao usuário definir qual formato é o mais adequado para aplicação das técnicas de mineração selecionadas, outro ponto importante é que baseado no tipo de conjunto de dados disponível para análise é que definem-se os padrões, relacionamentos ou informações que se consegue minerar. Nessa etapas todas as incoerências na base devem ser corrigidas e caso for necessário podem ser acrescentados mais atributos que sejam resultantes da combinação de outros ou que possam ser deduzidos de outros, como a idade que pode ser calculada a partir da data de nascimento, ou o total de atividades realizadas que pode ser efetuada por meio de uma soma.
3. Definição das Técnicas de MDE: A definição das técnicas é um passo importante, pois elas devem ser específicas para o tratamento da função de MDE estabelecida, as técnicas mais utilizadas para MDE são: a Estatística Descritiva, a Aprendizagem de Máquina e mais recentemente tem sido empregada a Aprendizagem Profunda, cabe salientar que cada uma dessas técnicas possui algoritmos, rotinas e/ou pro-

cedimentos específicos para manipulação dos dados. A literatura (AGGARWAL, 2015; BAKER; ISOTANI; CARVALHO, 2011; DE LOS REYES *et al.*, 2019; RIGO *et al.*, 2014), em muitos casos, não deixa claro as diferenças entre funcionalidades, funções e técnicas. Por exemplo, um problema que requer identificar alunos desistentes ou concluintes, para isso deve-se pensar em técnicas que possibilitem identificar esses padrões de comportamento nos alunos nessas duas categorias, isso seria a função, o propósito da MDE o de classificar esses alunos, entretanto também é possível entender como a técnica que seria aplicada, algoritmos de Aprendizagem de Máquina para classificação, nesse caso, por isso as vezes fica confusa a interpretação.

4. Delineamento de como essas técnicas serão aplicadas: Nessa etapa são selecionadas as ferramentas que vão dar suporte ao desenvolvimento de sistemas capazes de processar os dados, e gerar os resultados esperados.

Na Figura 04 pode-se observar as relações entre as etapas enunciadas, tendo em vista explicar a interatividade da função da MDE com as técnicas a serem utilizadas.

Figura 2. Etapas do processo de MDE



Fonte: Adaptado Aggarwal (2015) e Rometo e Ventura (2020)

A MDE, levando em consideração seus objetivos, técnicas e processo de funcionamento, formam uma importante metodologia para apoio a qualquer cenário educacional. Por meio da MDE, talvez seja possível acompanhar e compreender o processo de aprendizagem, bem como outros fatores que a influenciam. Como por exemplo, identificar que tipo de abordagem

instrucional (e.g. aprendizagem individual ou colaborativa) proporciona mais benefícios ao aluno, observando variáveis que representem seu engajamento com o curso. Além disso, abre-se a possibilidade de verificar se o aluno está aprendendo ou confuso, identificar níveis de motivação, envolvimento nas atividades on-line, descoberta de elementos ou indicadores comportamentais de conclusão e sucesso em um curso, identificar padrões de interação, descobrir estratégias que contribuam para a permanência dos estudantes (PURSEL *et al.*, 2016), bem como detectar possíveis fraudes, ou trapanças no sistema de aprendizagem. Estes fatores podem ajudar a personalizar o ambiente e os métodos de ensino, para oferecer melhores condições de aprendizagem (BAKER; ISOTANI; CARVALHO, 2011).

Em suma, a MDE tem se desenvolvido e pode ser considerada como uma das formas mais promissoras para extração de informações de bases de dados educacionais e suas técnicas têm se tornado cada vez mais eficientes e eficazes, graças ao número crescente de dados disponíveis e dos avanços computacionais. Para entender melhor como se deu esse desenvolvimento e como a MDE tem sido empregada no decorrer do tempo, são descritos os aspectos de sua evolução na sequência.

A evolução da Mineração de Dados Educacionais

A disponibilidade de grandes bases de dados educacionais, fomentada pelas modernas plataformas e mídias educacionais, combinadas com avanços na computação, formam a composição ideal para o surgimento da MDE. Embora existam relatos sobre publicações a respeito deste tema desde 1995 (ROMERO; VENTURA, 2007) o primeiro *workshop* foi realizado em 2005, em Pittsburgh, Pensilvânia, tendo sido seguido por várias oficinas e, em 2008, ocorreu a 1.^a Conferência Internacional sobre MDE realizada em Montreal, Quebec. As conferências anuais sobre MDE impulsionaram o surgimento do *Journal of Educational Data Mining*, que publicou sua primeira edição em 2009, na sequência, o primeiro manual de MDE foi publicado em 2010 (ROMERO *et al.*, 2010). Posteriormente, em 2011 a Sociedade Internacional de Mineração de Dados Educacionais foi formada com o objetivo de promover pesquisa científica na área interdisciplinar da MDE, organizando as conferências e os periódicos. No campo das publicações, uma primeira revisão de literatura foi apresentada por Romero e Ventura (2007), seguido de um modelo teórico proposto por Baker e Yacef (2009), e uma revisão bem mais abrangente sobre MDE foi desenvolvida por Romero e Ventura (2010). Na sequência, outras publicações iniciaram um amplo movimento de pesquisas nesse âmbito, e tiveram grande notoriedade.

Neste sentido, muitas publicações sobre MDE surgiram nos últimos anos, e algumas delas se dedicaram a implementar revisões de literatura sobre essa área. Tais estudos auxiliam no entendimento de como a MDE tem sido aplicada em diversos contextos educacionais,

seus objetivos, as técnicas mais utilizadas, verificação de resultados alcançados e validação dos benefícios proporcionados, identificação de avanços e também desafios que têm sido relatados por pesquisadores da área.

À vista disso, chamaram a atenção algumas publicações, que trouxeram grandes contribuições para pesquisadores interessados em MDE, quais sejam: Shahiri, Husain e Rashid (2015); Sukhija, Jindal e Aggarwal (2016); Schwendimann *et al.* (2017); Aldowah, Al-Samarraie e Fauzy (2019); e Romero e Ventura (2020), e foram fundamentais para um aperfeiçoamento e compreensão da evolução da MDE no decorrer de sua consolidação como área de pesquisa e são sintetizadas na sequência.

A primeira revisão sobre MDE analisada foi desenvolvida por Shahiri, Husain e Rashid (2015), essa forneceu uma visão geral das técnicas de mineração de dados que eram usadas para prever o desempenho dos alunos, em publicações datadas entre 2002 e 2015. O estudo também se concentrou em como os algoritmos de previsão poderiam ser usados para identificar os atributos mais importantes dentre a diversidade de dados dos alunos. Nessa revisão, os autores seguiram duas questões de pesquisa para estruturar os resultados: 1) Quais são os atributos mais importantes empregados na previsão do desempenho dos alunos; e 2) Quais as técnicas/algoritmos de previsão mais eficientes. Quanto aos principais atributos, Shahiri, Husain e Rashid (2015) apontam que foram usados com frequência a média cumulativa de notas e a avaliação interna usada por 10 dos 30 artigos selecionados para a revisão. Os autores também chegaram à conclusão, que a *Machine Learning* era a técnica mais usada e quanto à eficácia dos algoritmos as Redes Neurais tiveram a maior precisão (98%) para previsão do desempenho dos alunos, seguida das Árvores de Decisão (91%), depois as Maquinas de Vetores de Suporte e KNN com a mesma eficácia (83%), por fim, o método menos preciso foi o Naive Bayes (76%). Os autores afirmaram ainda, que prever o desempenho dos alunos é muito útil para ajudar educadores e alunos a melhorar o processo de ensino e aprendizagem. No entanto, é importante ressaltar que os índices de eficácia são resultado da interação entre a complexidade da questão de pesquisa com a qualidade (e algumas vezes, a extensão da base) dos dados, não sendo uma avaliação a respeito dos métodos em si. Aponta-se como ponto negativo da revisão, que algumas das pesquisas citadas pelos autores não apontam uma diversidade de métricas de avaliação dos algoritmos – uma informação importante, que deveria ser divulgada como parte dos resultados

Na sequência foi analisada a revisão sistemática desenvolvida por Sukhija, Jindal e Aggarwal (2016) que descreveram a evolução da MDE, trazendo à tona os aspectos e resultados de vários estudos divididos em 3 gerações: 1.^a geração de 2001 a 2005; 2.^a geração de 2006 a 2010; e 3.^a geração, de 2011 até 2015. No período de 2001 a 2005, as pesquisas se basearam no uso da MDE como uma ferramenta para antecipar os padrões que ajudam

na avaliação de cursos on-line. Dessa forma, os registros de dados e registros de atividades dos alunos foram usados para analisar seu comportamento. Os autores evidenciaram, que o início da literatura relacionada à MDE, produziu pesquisas com inclinação para o ambiente de aprendizagem baseado na *Web*, devido especialmente à grande disponibilidade de dados em cursos on-line. No final deste período, as pesquisas estavam com foco no uso de algoritmos evolutivos para mineração de dados da internet.

Referente ao período de 2006 a 2010, a MDE evoluiu e os estudos passaram a buscar a aplicação de algoritmos mais eficientes. Os bancos de dados usados se tornaram provenientes de sistemas de ensino à distância, baseados na *Web* e vinculados a grandes Instituições de Ensino, bem como o tamanho desses bancos de dados aumentou. Além disso, houve uma inclinação dos pesquisadores para análises preditivas de dados, com relação a prever os problemas e identificar os alunos em potencial, com alta probabilidade, de apresentar um desempenho acadêmico ruim, nesse sentido, sistemas de apoio à decisão para equilibrar a demanda e a oferta educacional também foram desenvolvidos. Durante esse período, a implementação da classificação baseada em Árvores de Decisão e Redes Neurais se acentuaram no contexto educacional, com diferentes objetivos. Ademais, citam que a pesquisa na área foi direcionada a dados relacionados ao ensino superior, bem como técnicas como *Online Analytical Processing* (OLAP³) em combinação com DELPHI⁴ foram muito utilizadas.

Finalmente, no que tange ao período de 2011 a 2015, a MDE evoluiu para incorporar técnicas melhores e mais eficientes, conseguindo integrar novas e mais eficientes regras de associações, ferramentas como WEKA⁵ começaram a ganhar popularidade, algumas pesquisas se voltaram a utilizar dados também do ensino médio e a comunidade de pesquisa trabalhou em função de uma aceitação comercial das técnicas de MD na educação. Os pesquisadores forneceram evidências para maior correlação entre diferentes construtos no sistema educacional, levando a uma maior anuência dos resultados obtidos entre estudantes e autoridades. Além disso, extensos estudos foram realizados na busca de encontrar uma solução para as altas taxas de abandono em diferentes contextos acadêmicos, nessa perspectiva também uma grande quantidade de estudos para estimar o desempenho dos estudantes, foram implementados. Sukhija, Jindal e Aggarwal (2016) ainda expuseram, que os bancos de dados usados neste período ficaram consideravelmente maiores do que os anteriores e essa crescente no volume dos dados foi acompanhada do desenvolvimento de novas técnicas para MDE, como por exemplo, a retomada dos estudos com *Deep Learning*.

3 OLAP é um conceito de interface com o usuário que proporciona a capacidade para manipular e analisar um grande volume de dados sob múltiplas perspectivas.

4 O Delphi é comumente conhecido como uma linguagem de programação. Mas, na verdade abrange um kit de desenvolvimento de software e ganhou propagação entre profissionais da área de desenvolvimento no segmento de aplicações de desktop, entretanto hoje é integrado e utilizado também para aplicações *Web* e *mobile*.

5 Weka é um Software livre do tipo *open source* para mineração de dados, desenvolvido em Java. Ao longo dos anos se consolidou como a ferramenta de data mining mais utilizada em ambientes acadêmicos.

Sukhija, Jindal e Aggarwal (2016) apontam cinco lacunas na área de MDE: (1) indisponibilidade de conjuntos de dados consistentes que sejam grandes o suficiente para refletir o sistema educacional e seu funcionamento; (2) necessidade de integração e versatilidade nos conjuntos de dados; (3) grande parte das técnicas de mineração foram aplicadas isoladamente e poucos trabalhos foram realizados utilizando técnicas híbridas; (4) havia falta de confiança das autoridades nos resultados da MDE e (5) necessidade de comparar métodos. Pode-se dizer que embora as descobertas dos autores fossem fortemente fundamentadas, o cenário se modificou bastante deste então.

No que se refere, à primeira lacuna ressalta-se que com a evolução dos cursos *e-learning*, bases com milhões de dados estão disponíveis, como exemplo, pode-se aludir ao trabalho desenvolvido por Northcutt, Ho e Chuang (2016), onde foi utilizada uma base de dados gerada a partir de uma plataforma de *Massive Open Online Course* (Curso Online Aberto e Massivo) com 1.893.092 de usuários, que produziram em média de 200 a 1500 interações com a plataforma, cada um, por curso realizado, portanto, a indisponibilidade de conjuntos de dados já não se configura mais como um problema. No que tange à segunda limitação pode-se dizer que em relação à integração das bases, ela se mantém, pois, não é possível integrar duas bases de forma simples, sem necessidade de um grande esforço de pré-processamento. Em relação à versatilidade dos dados – qualidade de não ser colinear, ou seja, dos dados não estarem relacionados – pode-se dizer que houve mudança, pois, vários tipos diferentes de dados são usados nos modelos.

A terceira lacuna, sobre uso de métodos híbridos, pode ser considerada a que mais não coincide com a realidade dos experimentos realizados na área de MDE atualmente, pois muitos pesquisadores têm empregado técnicas de MDE em conjunto com outras ferramentas de pesquisa como em Gallén e Caro (2017) que utilizaram algoritmos de Agrupamento e um Questionário respondido pelos alunos para analisar os motivos pelos quais uma pessoa se inscreve em um MOOC. Ademais, pode-se citar como exemplo o trabalho desenvolvido por Nen-Fu et al. (2018) que propõem um método para usar resultados de questionários e tipos de grupos predefinidos, por meio do algoritmo K-means, para classificar os alunos em MOOCs, com o intuito entender sua auto-organização nas fases iniciais de um curso, pois acreditam haver uma ligação entre a motivação e comportamento de aprendizagem. Com relação, à quarta limitação supõe-se que com os grandes avanços tecnológicos disponíveis e a consolidação da Inteligência Artificial (IA), presente no cotidiano das pessoas, a aceitação da MDE como aporte para tomada de decisões no setor acadêmico tenha crescido.

Enfim, quanto à quinta lacuna, é possível destacar que devido aos avanços tecnológicos muitos pesquisadores tem se dedicado a comparar novas técnicas de MDE com outras mais consolidadas, para verificação da eficácia. Nesse sentido, Gao et al. (2019) propuseram um

novo modelo para analisar o perfil de aprendizagem e o engajamento de alunos em MOOCs e para validar sua precisão os autores o compararam com algoritmos amplamente utilizados – Regressão Linear e Máquinas de Vetores de Suporte. Além desse exemplo, aponta-se a pesquisa de Waheed *et al.* (2020), que tinha como objetivo prever o abandono em MOOCs, utilizando um rede neural de Aprendizado Profundo, que foi comparada aos algoritmos de Regressão Logística e Máquinas de Vetores de Suporte.

A outra revisão incluída neste estudo foi elaborada por Schwendimann *et al.* (2016), que incluía, além de MDE Análise de Aprendizado (AA) e painéis de aprendizado. Os autores cunharam o termo “painéis de aprendizado” e o definiram da seguinte forma: “uma única exibição que agrega diferentes indicadores sobre aluno(s), processo(s) de aprendizagem e/ou contexto(s) de aprendizagem em uma ou várias visualizações”. Nesse sentido, os autores afirmaram que os painéis de aprendizado estão se tornando populares devido ao aumento do uso de tecnologias educacionais, como Sistemas de Gerenciamento de Aprendizagem (termo em inglês: *Learning Management Systems* - LMS) para a execução de diversos tipos de cursos em EAD como os MOOCs, que na opinião deles constituem a base para o desenvolvimento das áreas de AA e MDE.

A revisão de Schwendimann *et al.* (2016) foi realizada em 6 bases de dados: ACM Digital Library, IEEE Xplore, SpringerLink, Science Direct, Wiley e Google Scholar, retornando 346 artigos no total, dos quais 55 artigos foram incluídos na análise final. A revisão distinguiu entre 2 tipos de contribuições: Artigos que contribuíram com uma proposta teórica ou referencial (3 artigos; 5%); Artigos que descreveram a implementação de um painel de aprendizado (39 artigos; 71%), além de 13 artigos (5%) que apresentaram uma combinação dessas duas. A revisão finalizou delineando questões em aberto e futuras linhas sobre como trabalhar na área de painéis de aprendizado, indicando ainda que há uma necessidade longitudinal de pesquisas e de captação de dados em ambientes virtuais de aprendizado, assim como estudos que comparem sistematicamente designs de painéis diferentes.

A princípio o foco na MDE, pela revisão relatada, parece ficar difuso, entretanto no decorrer da análise percebe-se sua relação quando Schwendimann *et al.* (2016) citam os painéis de aprendizagem como subsídios para o desenvolvimento das áreas de MDE e AA. Tais painéis se constituem como uma forma sistemática de organização dos dados disponíveis dos alunos, para a finalidade de investigação. Os autores salientam ainda que os painéis apresentados nas publicações analisadas utilizaram principalmente os relatórios de Logs das atividades dos estudantes como fonte de dados, e algumas dessas pesquisas usaram Interface de Programação de Aplicativos (termo em inglês: *Application Programming Interface* – APIs) externas, bem como atividades escritas dos estudantes, e inclusive bancos de dados institucionais. Nesse sentido, a revisão de Schwendimann *et al.* (2016) auxiliou

a entender melhor quais tipos de dados podem servir ao propósito de examinar, por meio de MDE, o comportamento dos alunos, resultante da navegação em uma plataforma de ensino e aprendizagem.

Na sequência, a revisão de Aldowah, Al-Samarraie e Fauzy (2019) teve como foco o tema Mineração de Dados Educacionais e Análise de Aprendizado para o século XXI no ensino superior. Os autores relataram que as revisões anteriores sobre MDE e AA forneceram informações substanciais sobre a base teórica desse campo em rápido crescimento, no entanto, tais estudos não consideraram a associação entre diferentes técnicas de MDE e AA na resolução de problemas educacionais específicos e não produziram uma classificação clara das dimensões em que essas técnicas poderiam ser aplicadas com sucesso no ensino superior.

Aldowah, Al-Samarraie e Fauzy (2019) sugerem que essa revisão poderia ser aplicada como um guia para futuros estudos sobre o uso de técnicas de MDE e AA com o propósito de resolver problemas específicos de ensino e aprendizagem. Os autores conduziram a revisão com o intuito de responder a duas questões: “Como usar MDE e AA para resolver desafios práticos na educação?” e “Quais técnicas de mineração são mais adequadas para esses problemas?”. Para responder a essas perguntas, buscaram artigos das bases de periódicos e conferências: Scopus, Web of Science, Google Scholar, ERIC, Science Direct, DBLP, ACM Digital Library, IEEEXplore e Springer, encontrando 491 artigos publicados de 2000 até 2017.

Aldowah, Al-Samarraie e Fauzy (2019) afirmaram que as técnicas podem ser agrupadas em 4 grandes dimensões: Análise de Aprendizagem Suportada por Computador (*Computer-Supported Learning Analytics* – CSLA); Análise Preditiva Suportada por Computador (*Computer-Supported Predictive Analytics* – CSPA); Análise Comportamental Suportada por Computador (*Computer-Supported Behavioral Analytics* – CSBA); e Análise de Visualização Suportada por Computador (*Computer-Supported Visualization Analytics* – CSVA). Os resultados trazidos pelos autores podem ser sintetizados da seguinte forma: (1) As pesquisas sobre CSLA (120 artigos) concentraram-se principalmente no uso de análise estatística de dados para executar tarefas analíticas sofisticadas, a fim de investigar os comportamentos de aprendizagem colaborativa e de busca de informações dos alunos no contexto de um curso; (2) Os estudos sobre CSPA (253 artigos) focaram, em sua grande maioria, no uso de funções preditivas ou variáveis contínuas para sugerir maneiras eficazes de melhorar o aprendizado e o desempenho dos alunos, bem como avaliar a adequação do aprendizado; (3) As publicações sobre a dimensão CSBA (80 artigos), em maior parte, criaram modelos de comportamento, ações e conhecimento; (4) Os estudos sobre CSVA (38 artigos) concentraram-se em métodos para explorar visualmente os dados – usando gráficos interativos por

exemplo – para destacar informações úteis e produzir decisões precisas sobre as informações novas descobertas nos dados.

Os autores relataram que mineração sequencial de padrões, mineração de texto, mineração de correlação, detecção de outlier e mineração de estimativa de densidade, não são comumente usadas devido à complexidade na obtenção dos atributos necessários para regular ou adaptar-se às necessidades dos dados educacionais. Além disso, descobriram que na CSPA há uma taxa mais alta de tarefas de classificação, devido essa ser aceita como uma técnica eficaz para prever padrões de interesse e formar modelos de aprendizagem, promovendo tarefas específicas nesse sentido. Pode-se afirmar que essa revisão forneceu informações substanciais sobre a base teórica, metodológica e objetivos dessas áreas em expansão.

Finalizando a explanação sobre as revisões focadas em MDE e temas correlatos, que deram suporte ao desenvolvimento desse estudo, destaca-se Romero e Ventura (2020), que efetuaram um estudo sobre a MDE e AA, atualizando revisões anteriores (ROMERO; VENTURA, 2007, 2010, 2013, 2017). Partindo dessa premissa, a publicação de 2020 forneceu informações sobre o estado da arte, revisando as publicações da área no sentido de elucidar: os principais marcos; o ciclo de descoberta de conhecimento; os ambientes educacionais mais utilizados; as ferramentas específicas desenvolvidas; os conjuntos de dados disponíveis gratuitamente; os métodos e técnicas mais empregados; os principais objetivos; e por fim, as tendências futuras nessa área de pesquisa. Devido à grande amplitude da revisão, são abordados os itens considerados mais relevantes: as principais alterações na MDE e na AA e suas conclusões.

No que diz respeito as principais mudanças, Romero e Ventura (2020) afirmaram que na de 2010 a 2020 a MDE, como área de pesquisa, evoluiu enormemente e uma ampla gama de expressões relacionadas surgiram no estado da arte, os autores citam como principais as apresentadas no Quadro 1. Quanto às conclusões, Romero e Ventura (2020), apresentaram apontamentos relacionados aos seguintes aspectos: (1) A importância e evolução da MDE e AA; (2) Uma avaliação se as tendências encontradas na publicação de Romero e Ventura (2013) haviam se concretizado; e (4) tendências em MDE e AA.

Quadro 1. Expressões Relacionadas a MDE

EXPRESSÕES	DEFINIÇÃO
<i>A Academic Analytics (AA) e a Institutional Analytics (IA)</i>	Se preocupam com a coleta, análise e visualização de atividades do programa acadêmico como: cursos de graduação, cursos EAD, avaliação de cursos, alocação de recursos e gerenciamento para gerar <i>insights</i> institucionais. Portanto, estão focadas no desafio político/econômico.
<i>Teaching Analytics (TA)</i>	Refere-se à análise das atividades de ensino, aprendizagem, dados de desempenho, bem como do design, desenvolvimento e avaliação dessas atividades, está focada no desafio educacional do ponto de vista dos instrutores.
<i>Data-Driven Education (DDE) e Data-Driven Decision-Making in Education (DDDM)</i>	Referem-se a coletar e analisar sistematicamente vários tipos de dados educacionais, para orientar uma série de decisões com o intuito de ajudar a melhorar o sucesso de alunos e escolas.
<i>Big Data in Education (BDE)</i>	Denota à aplicação de <i>Big Data</i> – conotação básica resumida em volume, variedade e valor – para dados do ambiente educacional.
<i>Educational Data Science (EDS)</i>	Edefinida como o uso de dados coletados de ambientes/configurações educacionais para resolução de problemas nesse contexto. A ciência de dados é um conceito para unificar estatísticas, análise de dados, <i>Machine Learning</i> , métodos e técnicas relacionados.

Fonte: Romero e Ventura (2020)

No que se refere à importância da MDE e a AA, os autores salientaram que essas são duas comunidades interdisciplinares de cientistas da computação, cientistas de aprendizagem, psicometristas e pesquisadores de diversas áreas, mas todos com o mesmo objetivo, o de melhorar o aprendizado a partir dos dados. Os autores sugeriram que a área cresceu rapidamente nas últimas duas décadas, com duas conferências anuais⁶, dois periódicos específicos⁷ e com o aumento do número de livros, artigos, pesquisas e resenhas. Além disso, há uma corrente para mudar as pesquisas restritas a laboratórios para o mercado em geral, assim empregando a MDE e AA em instituições educacionais e escolas de todo o mundo. Os autores declararam, que em 2020 toda pesquisa educacional com relativa importância para o cenário envolve análise e mineração de dados. Isso indica que essas áreas se tornarão em breve maduras e amplamente utilizadas não apenas pelos pesquisadores, mas também por instrutores, administradores educacionais e negócios relacionados, em todo o mundo.

Com relação à comparação entre as tendências levantadas em sua pesquisa anterior (ROMERO; VENTURA, 2013) os autores destacaram que em 2013 dois direcionamentos foram levantados: o primeiro dizia respeito as ferramentas para MDE e AA, o qual ainda não foi completamente alcançado; e o segundo se referia ao desenvolvimento de uma cultura baseada em dados, que continua sendo um desafio (ROMERO; VENTURA, 2020). Referente à primeira tendência, de acordo com Romero e Ventura (2013), as ferramentas para MDE e AA deveriam ser disponibilizadas gratuitamente, para que uma população mais ampla pudesse utilizá-las, o que aconteceu, pois, uma grande variedade de ferramentas de finalidade específica está disponível de forma gratuita. Porém os autores ressaltaram que ainda é necessário desenvolver ferramentas para uso geral, que possam ser aplicadas em

6 Conferência Internacional sobre Learning Analytics & Knowledge (LAK) – Conferência Internacional sobre Educational Data Mining.

7 Journal of Educational Data Mining – Journal of Learning Analytics

várias tarefas e resolver diferentes problemas educacionais com a mesma interface/ferramenta, além disso, é necessário melhorar a portabilidade dos modelos obtidos por essas ferramentas (ROMERO; VENTURA, 2020). A segunda tendência pressupunha que educadores e instituições deveriam desenvolver uma cultura baseada em dados, utilizando-se deles para tomar decisões e melhorar seus processos de ensino, aprendizagem e administrativos. Todavia, de acordo com a pesquisa de 2020, a maioria das instituições e profissionais de ensino continuam cientes dos benefícios proporcionados pela análise de dados em larga escala, mas não adotaram essa cultura de forma efetiva nos seus processos gerenciais.

Romero e Ventura (2020) ainda sistematizaram os principais desafios relatados acerca de MDE e AA, quais sejam: transferibilidade e generalização; eficácia e aplicabilidade e interpretabilidade. Transferibilidade e Generalização se referem à utilização de modelos comuns para vários contextos, o que ainda não ocorre na prática. Eficácia e Aplicabilidade dizem respeito à realização de ações de intervenção a partir das análises de dados, assim há uma grande diversidade de modelos desenvolvidos os quais não se tem informações se foram ou são aplicados na prática das Instituições de Ensino e se são efetivos na resolução de problemas dessas instituições. Interpretabilidade refere-se à capacidade dos usuários compreenderem os modelos gerados para as análises de dados, acarretando um mau aproveitamento de tais modelos.

Finalmente, os autores propuseram algumas ideias, que disseram ser “visionárias e pessoais” que, em suas opiniões, podem formar tendências e direcionamentos muito promissores para a áreas de MDE e AA, para a década de 2020 resumidos no Quadro 2.

Quadro 2. Tendências para MDE e AA

TENDÊNCIA	ANÁLISE
<i>Levar em consideração todos os dados pessoais dos alunos durante toda a vida</i>	Atualmente, as informações consideradas na MDE e AA baseiam-se principalmente na interação de alunos com um único ambiente educacional, mas em um futuro próximo, graças ao grande volume de dados e à Internet das Coisas (IoT), pesquisadores serão capazes de ter informações disponíveis para cada aluno desde o nascimento até o momento e em tempo real. Isso implica a integração não apenas dos dados tradicionais de desempenho coletados das Instituições de Ensino e ambientes educacionais que cada aluno utilizou, mas também as informações sobre o status de cada aluno sob diferentes pontos de vista, como médico, familiar, econômico, religioso, sexual, relacionamento emocional, psicológico e assim por diante. Todos esses dados podem ser coletados a partir de várias fontes e elas poderiam ser fundidas para serem usadas com o intuito de melhorar e personalizar o processo de aprendizagem de cada aluno em cada momento específico de sua vida, o que suscitaria um novo nível de precisão.
<i>Aplicação e integração da MDE e AA aos futuros ambientes educacionais tecnológicos</i>	Na última década, os grandes avanços em tecnologias inovadoras permitiram o desenvolvimento de novos sistemas educacionais a partir de dispositivos móveis, a onipresença à realidade virtual, ambientes de realidade aumentada, hologramas e nas próximas décadas, satos quânticos serão associados à aplicação da Inteligência Artificial. Nesse contexto, não é errado pensar que os instrutores poderiam ser substituídos por máquinas sem que os alunos percebessem a mudança graças a avanços atuais em robôs humanoides inteligentes, agentes de conversação ou interfaces de voz e assistentes em ambientes educacionais. Mas esses sistemas precisam de técnicas de MDE e AA para analisar terabytes de dados e gerar modelos de análise portáteis em tempo real, a fim de enfrentar os desafios educacionais específicos desses futuros ambientes virtuais de aprendizagem.
<i>Análise e Mineração de Dados coletados diretamente do cérebro dos alunos para uma melhor compreensão do aprendizado</i>	O cérebro é o fator-chave para realmente entender como os alunos aprendem. Os avanços promissores na neurociência humana e neurotecnologia generalizada (interfaces cérebro-computador) estão dando origem a oportunidades sem precedentes de obter, coletar, compartilhar e manipular qualquer tipo de informação coletada do cérebro humano. Num futuro próximo, esses dados íntimos, sobre o estado psicológico e a atividade neural do aluno poderão ser analisados e minerados em tempo real, graças aos futuros dispositivos de eletroencefalografia de alta qualidade. Esses dados cerebrais, juntamente com outros dados multimodais poderiam ser integrados e usado pelos pesquisadores de MDE e AA, a fim de alcançar uma compreensão completa do processo de aprendizado.

Fonte: Adaptado Romero e Ventura (2020)

■ CONSIDERAÇÕES FINAIS

Em conclusão, com a análise e interpretação dessas publicações, que sistematizaram boa parte do estado da arte na área, em destaque dos últimos 20 anos, foi possível compreender o início do processo de adoção de MDE, quais eram as principais técnicas, os dados utilizados para formação de bancos/bases e os resultados que foram alcançados, enfim, de entender o modo pelo qual MDE se estabeleceu como uma área consolidada de pesquisa.

Além disso, explorar como ocorreu o processo de sua evolução, o qual foi constatado, teve início em decorrência fundamentalmente devido a dois fatores: 1) A adoção de grandes bases de dados na educação, impulsionado sobretudo pelo surgimento de cursos *e-learning*, como os do tipo MOOCs; e 2) O avanço das tecnologias computacionais, que são indispensáveis para aplicação das técnicas de MDE. Tais avanços, proporcionaram melhorias nas técnicas/ferramentas de MD já existentes, simplificando assim as tarefas dos pesquisadores e aperfeiçoando os resultados obtidos. Dessa forma, tais técnicas/ferramentas puderam ser aplicadas e testadas ao grande volume de dados educacionais disponíveis, consumando assim a MDE como um importante conjunto de técnicas e estratégia de pesquisa sobre dados educacionais.

Por fim, essas publicações propiciaram ainda, visualizar o cenário futuro para os próximos estudos com análises de dados educacionais. O que despertou interesse em descobrir quais seriam as principais tendências temáticas de pesquisas com MDE a contexto específicos da educação, como: no presencial, no ensino remoto, em cursos de Educação à Distância ou cursos *e-learning* diversos. Todavia, essa é uma proposta para um trabalho futuro, no qual pretende-se selecionar um contexto específico e analisar quais os propósitos da MDE, identificar tendências e tópicos de estudos mais proeminente, assim como encontrar quais as técnicas e algoritmos de MDE mais utilizados, além das principais oportunidades de pesquisa e os desafios mais relatados, formando um estudo amplo que apoie pesquisadores da área.

■ AGRADECIMENTOS

Agradeço ao Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul (IFRS) pela oportunidade de capacitação que possibilitou o desenvolvimento dessa pesquisa.

■ REFERÊNCIAS

1. AGGARWAL, Charu C. **Data Mining: The Textbook**. 1. ed. New York, USA: Springer, 2015. v. 1 *E-book*. Disponível em: <https://doi.org/10.1007/978-3-319-14142-8>
2. ALDOWAH, Hanan; AL-SAMARRAIE, Hosam; FAUZY, Wan Mohamad. Educational data mining and learning analytics for 21st century higher education: A review and synthesis. **Telematics and Informatics**, [S. l.], v. 37, p. 13–49, 2019. Disponível em: <https://doi.org/10.1016/j.tele.2019.01.007>
3. BAKER, Ryan; ISOTANI, Seiji; CARVALHO, Adriana. Mineração de Dados Educacionais: Oportunidades para o Brasil. **Revista Brasileira de Informática na Educação**, [S. l.], v. 19, n. 02, p. 3–13, 2011. Disponível em: <https://doi.org/10.5753/rbie.2011.19.02.03>
4. BAKER, Ryan S. J. D.; YACEF, Kalina. The State of Educational Data Mining in 2009 : A Review and Future Visions. **Journal of Educational Data Mining**, [S. l.], v. 1, n. 1, p. 3–17, 2009. Disponível em: <https://doi.org/10.5281/zenodo.3554657>
5. DE LOS REYES, Daniel A. Guimarães *et al.* Predição de sucesso acadêmico de estudantes: uma análise sobre a demanda por uma abordagem baseada em transfer learning. **Revista Brasileira de Informática na Educação**, [S. l.], v. 27, n. 1, p. 1–25, 2019. Disponível em: <https://doi.org/10.5753/rbie.2019.27.01.01>
6. EDM. **Educational Data Mining**. [s. l.], 2020. Disponível em: <http://educationaldatamining.org/>. Acesso em: 31 maio. 2020.
7. GALLEN, Rosa Cabedo; CARO, Edmundo Tovar. An exploratory analysis of why a person enrolls in a Massive Open Online Course within MOOCKnowledge data collection. *In*: 2017, Athens, Greece. **Global Engineering Education Conference, (EDUCON)**. Athens, Greece: IEEE, 2017. p. 1600–1605. Disponível em: <https://doi.org/10.1109/EDUCON.2017.7943062>
8. GAO, Lina *et al.* Modeling the effort and learning ability of students in MOOCs. **IEEE Access**, [S. l.], v. 7, p. 128035–128042, 2019. Disponível em: <https://doi.org/10.1109/ACCESS.2019.2937985>
9. NEN-FU, Huang *et al.* The Clustering Analysis System Based on Students' Motivation and Learning Behavior. *In*: 2018, **Proceedings of 2018 Learning With MOOCs, LWMOOCs 2018**. : IEEE, 2018. p. 117–119. Disponível em: <https://doi.org/10.1109/LWMOOCs.2018.8534611>
10. PURSEL, B. K. *et al.* Understanding MOOC students: Motivations and behaviours indicative of MOOC completion. **Journal of Computer Assisted Learning**, [S. l.], v. 32, n. 3, p. 202–217, 2016. Disponível em: <https://doi.org/10.1111/jcal.12131>
11. RIGO, Sandro José *et al.* Minerando Dados Educacionais com foco na evasão escolar: oportunidades, desafios e necessidades. **Revista Brasileira de Informática na Educação**, [S. l.], v. 22, n. 01, p. 168–177, 2014. Disponível em: <https://doi.org/10.5753/RBIE.2014.22.01.132>
12. ROMERO, C.; VENTURA, S. Educational data mining: A survey from 1995 to 2005. **Expert Systems with Applications**, [S. l.], v. 33, n. 1, p. 135–146, 2007. Disponível em: <https://doi.org/10.1016/j.eswa.2006.04.005>
13. ROMERO, Cristbal; VENTURA, Sebastin. Educational data mining: A review of the state of the art. **IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews**, [S. l.], v. 40, n. 6, p. 601–618, 2010. Disponível em: <https://doi.org/10.1109/TS-MCC.2010.2053532>

14. ROMERO, Cristóbal *et al.* **Handbook of Educational Data Mining**. 1. ed. Boca Raton, USA: CRC Press - Taylor & Francis, 2010. *E-book*. Disponível em: <https://doi.org/10.1201/b10274>
15. ROMERO, Cristobal; VENTURA, Sebastian. Data mining in education. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, [S. l.], v. 3, n. 1, p. 12–27, 2013. Disponível em: <https://doi.org/10.1002/widm.1075>
16. ROMERO, Cristobal; VENTURA, Sebastian. Educational data mining and learning analytics: An updated survey. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, [S. l.], v. 10, n. 3, p. 1–21, 2020. Disponível em: <https://doi.org/10.1002/widm.1355>
17. ROMERO, Cristóbal; VENTURA, Sebastián. Educational data science in massive open online courses. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, [S. l.], v. 7, n. 1, 2017. Disponível em: <https://doi.org/10.1002/widm.1187>
18. SCHWENDIMANN, Beat A. *et al.* Perceiving learning at a glance: A systematic literature review of learning dashboard research. **IEEE Transactions on Learning Technologies**, [S. l.], v. 10, n. 1, p. 30–41, 2017. Disponível em: <https://doi.org/10.1109/TLT.2016.2599522>
19. SHAHIRI, Amirah Mohamed; HUSAIN, Wahidah; RASHID, Nur'Aini Abdul. A Review on Predicting Student's Performance Using Data Mining Techniques. **Procedia Computer Science**, [S. l.], v. 72, p. 414–422, 2015. Disponível em: <https://doi.org/10.1016/j.procs.2015.12.157>
20. SUKHIJA, Karan; JINDAL, Manish; AGGARWAL, Naveen. The recent state of educational data mining: A survey and future visions. *In*: 2015, Amritsar, India. **3rd International Conference on MOOCs, Innovation and Technology in Education (MITE)**. Amritsar, India: IEEE, 2015. p. 354–359. Disponível em: <https://doi.org/10.1109/MITE.2015.7375344>
21. WAHEED, Hajra *et al.* Predicting academic performance of students from VLE big data using deep learning models. **Computers in Human Behavior**, [S. l.], v. 104, p. 1–13, 2020. Disponível em: <https://doi.org/10.1016/j.chb.2019.106189>