

MINERAÇÃO DE DADOS E BIG DATA NA EDUCAÇÃO

DATA MINING AND BIG DATE IN EDUCATION

THIAGO SETI PATRICIO

Mestrando em Mídia e Tecnologia da Universidade
Estadual Paulista – Unesp, Bauru, SP.
E-mail: thiago-2-pc@hotmail.com

MARIA DA GRAÇA MELLO MAGNONI

Professora doutora do Departamento de Educação da
Faculdade de Ciências (FC) da Universidade Estadual
Paulista – Unesp, Bauru, SP.
E-mail: sofia@fc.unesp.br

PATRICIO, Thiago Seti; MAGNONI, Maria da Graça Mello. Mineração de dados e big data na educação.
Revista GEMInIS, São Carlos, UFSCar, v. 9, n. 1, pp57-75, jan. / abr. 2018.

Enviado em: 01 de abril de 2018 / Aceito em: 05 de junho de 2018

RESUMO

O presente artigo apresenta uma análise de conceitos e utilidades que perpassam a Mineração de Dados (*Data Mining*) e *Big Data*, com o objetivo de demonstrar as consequências das aplicações desses dois pilares tecnológicos da atualidade no campo da educação. Nesta pesquisa é possível verificar que os autores fazem relatos de aplicações de *Data Mining* e *Big Data* na educação. Para isso, foram consultados livros, artigos científicos e reportagens que enfocam os conceitos propostos. O ponto central deste artigo é mostrar exemplos reais de aplicações de Mineração de Dados e *Big Data* na educação, bem como apresentar as vantagens de se armazenar grandes quantidades de dados e posteriormente tratá-los com a metodologia de Busca de Conhecimento em Banco de Dados (KDD). Por exemplo, em busca de melhor compreender o comportamento dos estudantes e a maneira como os mesmos aprendem, o trabalho efetuado por pesquisadores em Mineração de Dados educacionais tem investido no uso e na melhoria de técnicas de mineração de dados para obter conhecimentos úteis a partir desses dados. Tais conhecimentos podem servir de apoio para a melhoria das práticas em educação à distância ou presencial, além de ser uma importante ferramenta para viabilizar a personalização do ensino.

Palavras-chave: Mineração de Dados. *Big Data*. Dados. Educação.

ABSTRACT

This article presents an analysis of concepts and utilities that go through Data Mining and Big Data, in order to demonstrate the consequences of the applications of these two technological pillars of the present time in the field of education. In this research it is possible to verify that the authors report on Data Mining and Big Data applications in education. For this, were consulted books, scientific articles and reports that focus on the proposed concepts. The central point of this article is to show real examples of Data Mining and Big Data applications in education, as well as to present the advantages of storing large amounts of data and later treat them with the Database Knowledge Search methodology (KDD). For example, in order to better understand students' behavior and the way they learn, the work done by researchers in Educational Data Mining has invested in the use and improvement of data mining techniques to gain useful insights from these data. Such knowledge can support the improvement of practices in distance or face-to-face education, as well as being an important tool to enable personalized teaching.

Keywords: Data Mining. Big data. Data. Education.

INTRODUÇÃO

Com as Tecnologias da Informação e Comunicação (TICs)¹ emergentes, na contemporaneidade, e o crescimento exponencial de dados gerados na Internet, conceitos como Mineração de Dados (*Data Mining*) e *Big Data* (grandes dados, em tradução livre) são cada vez mais vistos na literatura de Tecnologia da Informação (TI), bem como em estudos de caso e pesquisas que buscam adequar estes conceitos de forma multidisciplinar, como por exemplo, nos negócios, na educação, na sociedade, no varejo, na saúde, na indústria, entre outros.

Em 2007, se toda informação gerada globalmente fosse armazenada em CD-Roms, e se estes fossem empilhados, a pilha ultrapassaria a distância existente entre a Terra e a Lua. Logo, essa quantidade de informações devidamente armazenada já supera a que existia na famosa biblioteca de Alexandria, que fora fundada em 300 a.C e possuía um dos maiores acervos do mundo antigo. (MOUTINHO, 2011)

No tocante ao crescimento cada vez maior de dados, faz-se necessário a aplicação de técnicas e algoritmos que visam investigar esses dados em busca de informações relevantes em repositórios, principalmente no que tange a tomada de decisões, e entre essas técnicas de investigação encontra-se a Mineração de Dados e o *Big Data*.

De acordo com Côrtes, Porcaro e Lifschitz (2002), a Mineração de dados constitui-se como uma etapa de um processo maior de descoberta de informações denominado de *Knowledge Discovery in Database (KDD)*, ou Busca de Conhecimento em Banco de Dados. Para mais, o KDD representa uma metodologia de descoberta de informações a partir de uma sequência definida de passos, sendo eles: Limpeza de dados; Integração dos dados; Seleção dos dados; Transformação dos dados; Mineração dos dados; Avaliação dos Padrões e Apresentação e assimilação do conhecimento. (HAN; KAMBER, PEI, 2012)

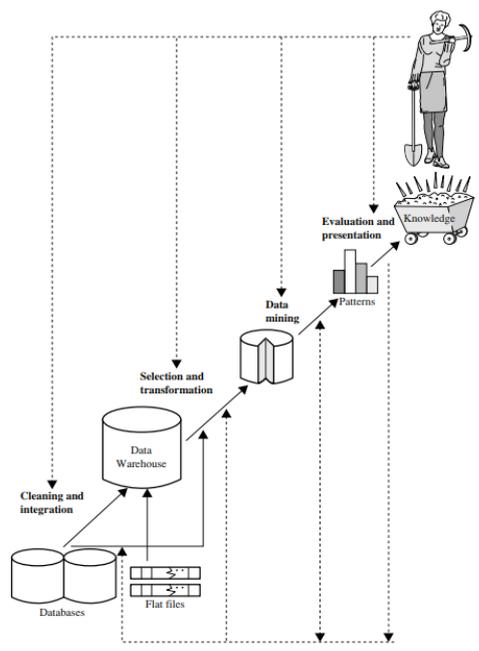
Ademais, a Mineração de Dados procura identificar relações entre dados,

¹ “Textos, sons e imagens são transformados em *bits*, acarretando a substituição das diferentes tecnologias que eram necessárias para as várias transmissões analógicas (telégrafo para texto, telefonia para voz, radiodifusão para sons e imagens, etc.), por redes integradas de usos múltiplos (via cabo ótico, satélites ou radiodigitais).” (OLIVEIRA, 2003, p. 85, grifo do autor).

e a partir desse processo, criar informações que podem gerar novos conhecimentos para o desenvolvimento da ciência, assim como também para tomada de decisões para subsidiar novas ações e transformar a realidade atual. (BARBOSA; CARVALHO; ANDRADE, 2017)

Na Figura 1 nota-se claramente que a Mineração de Dados é apenas uma etapa no processo de Descoberta de conhecimento em bases de dados.

Figura 1 – Processo de KDD.



Fonte: Han, Kamber e Pei (2012).

Logo, dada uma determinada base de dados, a etapa de Mineração de Dados é a responsável pela aplicação de algoritmos que irão fazer a extração de dados, e posteriormente enviar os padrões encontrados para a etapa de Validação e Apresentação, onde os mesmos serão então convertidos em informações a serem validadas pelo usuário.

Dentre as ferramentas disponíveis para Mineração de Dados, uma das que mais se destaca é a *Waikato Environment for Knowledge Analysis (WEKA)*, que de acordo com Abernethy (2010) trata-se de um *software* de código aberto e gratuito que possibilita a transformação de dados em conhecimento útil.

Outra ferramenta voltada para fins acadêmicos é a *RisingMiner*, que é uma ferramenta *web* para mineração de regras de associação baseada no algoritmo *Fuzzy Ontology Generalized Association Rules (FOntGAR)*². A ferramenta permite ao usuário entrar com parâmetros a fim de obter regras customizadas por meio de taxonomias,

² Algoritmo de mineração de regras de associação generalizadas desenvolvido por Rodrigo Moura Juvenil Ayres, como parte da dissertação “Mineração de Regras de Associação Generalizadas Utilizando Ontologias Difusas e Similaridade Baseada em Contexto”, de 2012.

que são estruturas de classificação para recuperar e organizar informações hierarquicamente. (PATRICIO, 2014)

O Quadro 1 mostra as seis melhores ferramentas de Mineração de Dados de código aberto.

Quadro 1 – As seis melhores ferramentas de mineração de dados de código aberto.

Ferramenta	URL
RAPIDMINER	http://rapidminer.com/products/studio/
ORANGE	http://orange.biolab.si/
KNIME	http://www.knime.org/
WEKA	http://www.cs.waikato.ac.nz/ml/weka/
KEEL	http://www.keel.es/
R	http://www.r-project.org

Fonte: Rangra e Bansal, 2014. (Modificado pelo autor)

Apesar das aplicações de Mineração de Dados serem implementadas nos mais diversos setores, como saúde, varejo, indústria, serviços financeiros, entre outros, o foco deste trabalho é o setor de educação, onde a Mineração de Dados atua no sentido de desenvolver e adaptar algoritmos já existentes para compreender melhor os dados oriundos de contextos educacionais, produzidos por alunos e professores, e considerando Sistemas Tutores Inteligentes (STIs), Ambiente Virtual de Aprendizagem (AVA), entre outros.

Segundo Costa et al. (2012), a utilização crescente de AVA no processo de ensino-aprendizagem proporciona a geração de um grande volume de dados, que envolve interações dos alunos e docentes com os ambientes de aprendizagem. Os autores ainda destacam que neste tocante pode-se citar o caso da Universidade Aberta do Brasil³, onde são gerados montantes consideráveis de dados por meio do uso de AVAs, e a Mineração de Dados poderia ser de grande benefício no que concerne a análise apropriada desses dados para a geração de informação útil.

Outro conceito importante no tocante a dados e informações é o de *Big Data*, cujo significado contemporâneo mais comum alude a um massivo conjunto de dados armazenados, que podem advir da Internet, dos mais variados dispositivos tecnológicos

³ <http://www.capes.gov.br/uab>

como *Smartphones, Notebooks, Tablets, Computador Pessoal (PC), Internet das Coisas* – trata de dispositivos como TVs, geladeiras, carros, relógios, óculos, e demais objetos do cotidiano que possuem a capacidade de estarem conectados à Internet; entre outros.

Para Enomura (2014), a definição do termo *Big Data* como se conhece atualmente surgiu em 2001 com Doug Laney, da empresa Gartner Group⁴, que na ocasião propôs que o conceito deveria ser sustentado sobre o tripé: Volume, Velocidade e Variedade. Outrossim, a autora ainda salienta que pode-se entender o *Big Data* como:

[...] bancos de dados tão grandes, que precisam de formas inovadoras de processamento de informação para uma melhor percepção e tomada de decisão. Na prática, significa que ferramentas poderosas estão minerando informações com mais eficiência para fazer coisas como: mapear o trânsito para sugerir melhores rotas, monitorar interesses para oferecer serviços e produtos personalizados, e até acelerar a decodificação do genoma humano.

Nesse contexto, os dados podem ser divididos em estruturados e não estruturados. O primeiro sendo informações fornecidas organizada-mente, como um levantamento de perfil dos clientes ou uma pesquisa eleitoral do Ibope; já os não estruturados são aqueles coletados num “caos organizado”, como comentários no Facebook, movimentações do cartão de crédito e pesquisas no Google.

Dessarte, especialistas e cientistas de dados já caracterizam o *Big Data* como uma junção de 5 Vs, a saber: Volume, Velocidade, Variedade, Veracidade e Valor. Ademais, o volume é a característica mais reconhecida neste conceito, e diz respeito ao tamanho do repositório de dados e dos arquivos armazenados, bem como a capacidade de processar esses dados, que é denominada de *Big Data Analytics*. A velocidade está ligada ao tempo de processamento (tempo real ou não) dos dados no *Big Data*. A variedade está interligada as categorias de dados que podem advir de sistemas estruturados e não estruturados (e-mail, Facebook, etc). A veracidade diz respeito à autenticidade e sentido dos dados. O valor é o retorno do investimento que deve existir quando uma solução de *Big Data* é implantada. (ISHWARAPPA; ANURADHA, 2015)

No que concerne ao *Big Data* no setor de educação, pode-se afirmar que sua utilização é deveras importante para capturar informações sobre os alunos, assim como suas interações com os ambientes de aprendizagem, conteúdos educacionais, avaliações, e assim favorecer para que haja melhorias no processo de aprendizagem, auxiliando gestores, educadores e formuladores de políticas educacionais. (SCAICO; QUEIROZ; SCAICO, 2014)

Para Scaico, Queiroz e Scaico (2014):

⁴ Site da empresa: <https://www.gartner.com/en>

O uso de *big data* pode reunir e analisar uma imensidade de dados que são produzidos pelas relações de ensino-aprendizagem. Com o uso de tecnologias pode-se coletar muito mais informações, em variados graus de granularidade, que os métodos manuais não são capazes de coletar. Capturar evidências que apontem a sequência de passos que um estudante usou para resolver um problema, suas estratégias, o número de conselhos que precisou receber e o tempo gasto na análise de cada parte do problema são alguns exemplos. A capacidade de processar massas de dados em escala, através da análise e da comparação de comportamento de milhares de estudantes, é muito importante para gerar conhecimentos generalizáveis sobre o processo de aprender e as dificuldades inerentes a determinados conteúdos ou estratégias pedagógicas.

Como adendo, pode-se assinalar que todo esse processo visa exclusivamente uma alteração da práxis social no tocante a educação, visto que não apenas o cotidiano escolar é modificado, mas a prática de se educar passa a considerar a cultura e o saber de cada aluno no processo educativo, tornando o docente conhecedor das características de cada aluno em particular, assim como suas dificuldades e nível de conhecimento, para enfim refletir e pôr em prática o desenvolvimento de um pensamento pedagógico com base na adaptação e personalização.

Nesse sentido, Pavan (2016) colabora:

O cotidiano escolar, onde ocorre o desenvolvimento da aprendizagem também permite uma afetiva reconstrução do conhecimento quando se dá espaço para a integração entre as diferentes maneiras de expressão, abrindo possibilidades de a escola reconhecer efetivamente o aluno como sujeito da reconstrução do conhecimento. Todos os conceitos educacionais são renovados a cada geração, visto que em cada momento e época histórica os indivíduos têm uma determinada visão de mundo, bem como um conjunto de conhecimentos e de valores. Essa renovação não quer dizer que se esqueça o velho conceito, mas que esteja considerando as novas maneiras de viver que caracteriza um grupo ou sociedade.

MATERIAL E MÉTODOS

Para Prodanov e Freitas (2013):

A pesquisa científica é uma atividade humana, cujo objetivo é conhecer e explicar os fenômenos, fornecendo respostas às questões significativas para a compreensão da natureza. Para essa tarefa, o pesquisador utiliza o conhecimento anterior acumulado e manipula cuidadosamente os diferentes métodos e técnicas para obter resultado pertinente às suas indagações.

Do ponto de vista de sua natureza, a pesquisa adotada para a confecção deste artigo é a chamada pesquisa básica, que de acordo com Prodanov e Freitas (2013), trata-se do envolvimento de “verdades e interesses universais”, e possui como objetivo o desenvolvimento da ciência, sem o envolvimento de uma aplicação prática prevista.

A pesquisa também lança mão de investigação da literatura de Mineração de Dados e *Big Data* através de referencial teórico, especialmente livros, artigos científicos e reportagens a fim de embasar teoricamente o trabalho.

Por adendo, a pesquisa bibliográfica é deveras importante no sentido de fornecer dados atuais e relevantes que estão relacionados com o tema de pesquisa proposto. Logo, ao lançar mão desta técnica de levantamento de dados, o pesquisador faz uma espécie de apanhado geral da literatura já existente sobre os conceitos trabalhados, a fim de descobrir no material coletado uma fonte rica de informações, bem como também de subsídios para o trabalho. (MARCONI; LAKATOS, 2003)

Segundo Prodanov e Freitas (2013), a pesquisa bibliográfica é aquela que é:

[...] elaborada a partir de material já publicado, constituído principalmente de: livros, revistas, publicações em periódicos e artigos científicos, jornais, boletins, monografias, dissertações, teses, material cartográfico, internet, com o objetivo de colocar o pesquisador em contato direto com todo material já escrito sobre o assunto da pesquisa. Em relação aos dados coletados na internet, devemos atentar à confiabilidade e fidelidade das fontes consultadas eletronicamente. Na pesquisa bibliográfica, é importante que o pesquisador verifique a veracidade dos dados obtidos, observando as possíveis incoerências ou contradições que as obras possam apresentar.

RESULTADOS E DISCUSSÕES

A penetração de tecnologias como *Big Data* e *Data Mining* na educação encontram sentido principalmente pelo exponencial crescimento de sistemas de gestão do aprendizado, que visam coletar os dados dos alunos. Nos dias atuais percebe-se que cada vez mais os estudantes utilizam telefones celulares, em especial os *smartphones*, que são dão acesso a inúmeros aplicativos e a Internet. Dessarte, toda atividade do aluno dentro ou fora da instituição de ensino acarreta em uma massiva quantidade de dados, que são desperdiçados se não forem utilizados, logo, faz-se necessário a aplicação de tecnologias de *Big Data* como a Mineração de Dados para a análise da aprendizagem no contexto educacional. (SIN; MUTHU, 2015)

Para Sin e Muthu (2015), à medida que crescem os dados envolvidos no campo da educação, torna-se imprescindível a utilização de técnicas de *Big Data* para minerar

esses dados e transformá-los em informação pertinente. Como adendo, os autores ainda destacam que, ao analisar a literatura acerca de *Big Data* e *Data Mining* na educação, pôde-se notar que os trabalhos enfocam principalmente conceitos de introdução a Mineração de Dados educacionais, bem como a predição ou medição do desempenho do estudante utilizando técnicas de *Data Mining*.

Grinbergas (2015) explica que as técnicas de análise de *Big Data* podem ser utilizadas em diversas áreas, como o varejo, mercado financeiro, setor de tecnologia, entre outros, e acrescenta que estas técnicas também podem ser de grande valia para a gestão de instituições educacionais e otimização do desempenho dos estudantes, visto que:

“O volume e a velocidade de dados gerados vêm atingindo níveis sem precedentes. Para se ter uma ideia da dimensão disso, 90% dos dados disponíveis no mundo foram gerados nos últimos dois anos”, afirma Rachel Horta, CEO da Hekima, empresa que faz análise de Big Data. Na escola, a coleta de dados é feita por professores e gestores em situações como a atribuição de notas, a realização dos cálculos de frequência, o preenchimento do Censo e a aplicação de avaliações padronizadas. “Em um primeiro momento, reunir esses indicadores servia para prestar contas à sociedade e mostrar índices de desempenho. A ideia se encerrava aí. Entretanto, essas informações são preciosas e começam a ser entendidas como base para uma avaliação diagnóstica do ensino e para a tomada de decisões”, comenta Luis Antonio Laurelli, diretor executivo da Tuneduc, empresa de Big Data voltada ao universo educacional. (GRINBERGAS, 2015)

No Brasil, umas das instituições de ensino que lançam mão de técnicas de *Big Data* é o Colégio Dante Alighieri, que de acordo com a coordenadora geral de tecnologia do colégio Valdenice Minatel Melo de Cerqueira: “A proposta é dar instrumentos e números consistentes ao gestor, para que possa tomar as decisões acertadas”, sendo que para tal não se faz mais necessário aguardar notas e provas dos alunos que apontem se os mesmos estão assimilando corretamente o conteúdo ou não, pois de acordo com a coordenadora: “O professor já consegue preparar suas próximas aulas com base no *feedback* que teve em sala. Por exemplo: no fim de uma aula, ele lança um quiz e rapidamente consegue avaliar o quanto daquele conceito foi adquirido. Na seguinte, já estará preparado para reforçar o que foi a maior dificuldade em cada turma”. (GRINBERGAS, 2015)

Ademais:

Além de informações mais generalizadas, é possível ainda analisar alunos individualmente. “Em vez de recomendar um mesmo conteúdo para todos de uma turma, é possível fazer indicações específicas

para cada um. Queremos cada vez mais personalizar a experiência do estudante e o que nos permite fazer isso são os dados que temos sobre ele”, explica Leandro Faria, CEO da Datastorm, empresa do grupo Stefanini especializada em soluções de Big Data e analytics. O Colégio Santa Marcelina, de Belo Horizonte, é um que está aplicando o conceito nas aulas de matemática e redação do 7º ano do ensino fundamental e nas do 3º do ensino médio. “Por meio dos relatórios oferecidos, o professor personaliza sua intervenção, indicando atividades, videoaulas, oficinas de produção de texto de acordo com as necessidades e possibilidades de cada estudante”, diz Rita Rangel, coordenadora pedagógica da instituição. (GRINBERGAS, 2015)

Na contemporaneidade, a Mineração de Dados em contextos educacionais tem se tornado um campo emergente de exploração de dados em ambientes educacionais, onde técnicas e algoritmos de *Data Mining* são aplicados com o intuito de gerar novos conhecimentos voltados ao processo de ensino-aprendizagem, onde faculdades e instituições de ensino heterogêneas, porém afiliadas à mesma universidade, possam implementar um modelo unificado de planejamento acadêmico com o intuito de otimizar a utilização dos recursos disponíveis. (JINDAL; BORAH, 2013)

De acordo com Jindal e Borah (2013), o processo de Mineração de Dados educacionais envolve grupos diferentes de pessoas interessadas (*stakeholders*), que participam direta e indiretamente do contexto referido, tais grupos podem ser classificados como: Grupo primário (engloba os participantes de forma direta no contexto educacional, como estudantes e professores); Grupo secundário (agrupamento de pessoas indiretamente envolvidas com o setor educacional, como ex-alunos e pais de alunos); e por fim o Grupo híbrido (constituído por pessoas envolvidas no processo administrativo, como empregadores, planejadores educacionais e especialistas).

Atualmente, a quantidade de universidades e estudantes que nelas ingressam estão aumentando cada dia mais, portanto, faz-se necessário apoiar-se em tecnologias de Mineração de Dados a fim de otimizar o padrão educacional, maximizando o sistema de ensino com algoritmos de *Data Mining*, que visam descobrir informações importantes que ajudem a instituição a melhorar a taxa de candidatos bem-sucedidos, a questão da proporção de abandono escolar, entre outros. (NITHYA; UMAMAHESWARI; UMADEVI, 2016)

Ademais, com o crescimento nas últimas décadas de universidades, tanto do setor público como do privado, houve também uma crescente procura por parte dos alunos em cursos de graduação e pós-graduação, o que traz à tona a importância de se conhecer os requisitos de cada estudante, bem como sua progressão acadêmica. Outrossim, a Mineração de Dados no contexto educacional pode auxiliar a responder

essas questões de previsão de perfis dos alunos e seu desempenho, e *stakeholders* de educação. (SRIVASTAVA; SRIVASTAVA, 2013)

Para Srivastava e Srivastava (2013) a aplicação de Mineração de Dados no contexto educacional é diferente no tocante a estudantes e professores. Para os estudantes, o foco de *Data Mining* é investigar por meio de algoritmos atividades, tarefas e recursos de aprendizagem que os alunos acessam com base em suas atitudes e gosto pessoal, já para os professores, o intuito é obter maior retorno por parte dos alunos para poderem estruturar melhor suas aulas e o curso em questão, e classificar cada aluno de acordo com suas necessidades, para enfim melhorar as atividades do curso e adaptar melhor os conteúdos propostos.

De acordo com Dias (2017):

A tomada de decisão baseada na análise dos dados é bastante útil. Como exemplo, um projeto para uma empresa energética, que teve como objetivo aplicar técnicas de Data Mining para criação de modelos que façam a previsão de demanda de consumo de energia elétrica por regiões. Para tanto, foram utilizados registros de consumo de energia elétrica ao longo de períodos anteriores. Equivalentemente, dados de alunos podem ser minerados para que seja feita a conferência entre a aplicação pedagógica e o aprendizado do aluno. Com base nessa informação o professor teria a possibilidade de entender se sua abordagem está de fato contribuindo com o aluno.

As tarefas de Mineração de Dados na educação podem ser de grande valia tanto para previsões, classificação, associações e agrupamento, lançando mão de algoritmos que executam essas tarefas. Como adendo, um algoritmo de previsão, por exemplo, pode auxiliar a instituição a prever o desempenho do aluno com base no seu comportamento atual, algoritmos de classificação podem classificar os alunos em taxonomias de acordo com as necessidades de cada um, algoritmos de associação podem associar o aluno com algum conteúdo ou curso específico de acordo com suas preferências. Srivastava e Srivastava (2013) destacam que o papel das tarefas de *Data Mining* em educação é principalmente adaptar o ensino ao aluno através de recomendações para que o mesmo melhore sua aprendizagem, e até mesmo na criação de um ambiente de aprendizagem individual.

Cumprе assinalar que apesar do processo de ensino personalizado ser hodierno, este fundamenta-se em uma ideia antiga, visto que em tempos de outrora, o grande educador Paulo Freire já explicitava a necessidade de valorizar-se a cultura de cada educando no processo de aprendizagem. Dessarte, Freire acreditava que o enfoque na individualidade de cada pessoa era uma possível solução para o aperfeiçoamento do ensino. (HEKIMA, 2016)

Por adendo, Freire:

[...] jamais concordou com práticas educacionais que transmitissem aos sujeitos um saber já construído. Ele acreditava que o ato de educar deve contemplar o pensar e o concluir, contrapondo a simples reprodução de ideias impostas para ele alfabetização deveria ser sinônima de reflexão, argumentação e criticidade. (DREYER, 2011)

No tocante à *Data Mining* acrescenta-se que:

O âmbito da mineração de dados educacionais cria e ajusta técnicas e algoritmos utilizados atualmente, ao passo de compreender e contextualizar para a educação, considerando as plataformas de interação, tais como Ambientes Virtuais de Aprendizagem (AVAs), Redes Sociais, Sistemas Tutores Inteligentes (STIs). Com essa metodologia, pretende-se conhecer melhor o estudante em sua aprendizagem. (DIAS, 2017)

Em um estudo de caso realizado em 2016, o pesquisador Amjad Abu Saa, do Departamento de Tecnologia da Informação da Universidade de Ciência e Tecnologia de Ajman, nos Emirados Árabes Unidos, conduziu uma pesquisa onde era descobrir as relações entre os fatores pessoais e sociais dos alunos, e seu desempenho no semestre anterior utilizando técnicas de Mineração de Dados como a Classificação.

Ademais, o estudo foi realizado com os dados amostrais de um grupo de estudantes da Universidade de Ciência e Tecnologia de Ajman, lançando mão também de pesquisa online confeccionada nos Formulários Google⁵, sendo que os dados foram coletados de forma anônima, totalizando ao final da pesquisa em um conjunto de dados de 270 registros. Toda a implementação e processamento de *Data Mining* foi efetuado utilizando as ferramentas RapidMiner⁶ e WEKA⁷. (SAA, 2016)

O estudo de Saas (2016) pôde mostrar que a aplicação de Mineração de Dados foi capaz de prever alguns cenários e padrões, como a propensão de alunos com notas maiores em receber descontos da universidade, alunos com mães que trabalham recebem notas superiores, estudantes do sexo feminino tem a probabilidade de obterem notas mais altas que os estudantes do sexo masculino.

Seguidamente, em outra pesquisa realizada por Divna Krpan e Slavomir Stankov, da Faculdade de Ciência na Universidade de Split, na Croácia, é vista a aplicação de *Data Mining* em dados coletados dos alunos no ambiente de aprendizado

5 Disponível em: <https://www.google.com/forms/about/>

6 Ferramenta RapidMiner, disponível em: <https://community.rapidminer.com/t5/Getting-Started-Forum/RapidMiner-7-3-is-now-available-for-download/td-p/34552>

7 Ferramenta WEKA, disponível em: <https://www.cs.waikato.ac.nz/~ml/weka/downloading.html>

*Modular Object-Oriented Dynamic Learning Environment (MOODLE)*⁸.

No estudo foram verificados os dados de um grupo experimental de 52 alunos que durante três semanas utilizaram o sistema Moodle para aprendizado, sendo 8 testes online de múltipla escolha foram realizados no Moodle. Os dados coletados para cada aluno foram: tempo gasto para execução do teste, pedido de *feedback* para correção das tarefas, *feedback* recebido e se os estudantes verificaram o *feedback*. (KRPAN; STANKOV, 2012)

Os testes demonstraram que parte dos alunos fizeram lições em casa através do sistema, mas com o desempenho baixo, e outra parte, que não possui computador em casa, e foi monitorada dentro da faculdade, não responderam bem ao ambiente controlado, e ficaram apenas navegando e procrastinando pelas páginas das lições. (KRPAN; STANKOV, 2012)

Logo, pode-se dizer que sistemas de ensino a distância possuem um grande potencial para a aplicação de *Data Mining*, pois:

Os dados obtidos em ambientes de EaD, como grandes quantidades de texto envolvendo discussões (síncronas e assíncronas) em chats, fóruns de discussão, wikis, blogs, e outras formas de interação textual entre estudante-estudante e estudante-professor, têm excelente potencial de serem utilizadas para realizar-se mineração de texto de descobrir modelos interessantes sobre os alunos. (BAKER; ISOTANI; CARVALHO, 2011)

Dessarte, a emergência em torno da utilização de *Big Data* e *Data Mining* nos mais diversos setores e áreas da sociedade, tem feito crescer a demanda por profissionais que saibam lidar com este grande volume de dados para gerar informações de valor para seus negócios, assim como área de pesquisa e atuação. Logo, para responder as expectativas em torno de profissionais e pesquisadores especializados em dados, é que algumas renomadas universidades do Brasil possuem cursos em seu catálogo voltados para *Big Data* e *Data Mining*.

A Fundação Getúlio Vargas (FGV), por exemplo, oferece o curso de Especialização em Big Data Analytics, cujo objetivo é mostrar as principais tecnologias de análise de dados no *Big Data* disponíveis no mercado, com enfoque na importância de profissionais que saibam lidar com análise de dados. A Universidade Presbiteriana Mackenzie

⁸ O sistema Moodle é considerado uma plataforma de aprendizado à distância, em especial no que tange ao auxílio de docentes na criação de aulas e cursos online, na maioria das vezes em caráter de suporte ao ensino presencial. Cumpre assinalar que o Moodle baseia-se na filosofia educacional do construcionismo, cujo cerne teórico apresenta a construção do conhecimento partindo da mente do educando, em contrapartida ao modelo tradicional instrucional, onde o aluno recebe a transmissão do saber por meio de livros e aulas expositivas sem foco na interatividade. (SABBATINI, 2007)

oferta por sua vez o curso de Pós-Graduação em Ciência de Dados (Big Data/Analytics), que foca em profissionais de análise de dados. A Universidade de Taubaté (UNITAU) dispõe do curso de Especialização em Gestão de Projetos Business Intelligence, cujo objetivo é apresentar técnicas de inteligência de negócios para pesquisa e indústria, assim como softwares próprios para Mineração de Dados. (MERKER, 2014)

Por adendo, a Faculdade BandTec disponibiliza o curso de Pós-Graduação em Big Data & Analytics, com enfoque em governança de dados, visão analítica, e tomada de decisões baseada em fatos. A Faculdade e Tecnologia FIAP oferece o curso de MBA em Big Data, que visa preparar profissionais a modelar dados não-estruturados, na exploração de dados, assim como na compreensão de implicações de segurança e de ética quanto às aplicações de *Big Data*. (MERKER, 2014)

CONCLUSÃO

Na atualidade, pode-se dizer que por mais que dados sejam gerados em quantidades cada vez mais exponenciais, esses mesmos dados sem conhecimento pertinente são inúteis. Logo, o *Big Data* e a Mineração de Dados são essenciais no armazenamento, limpeza, extração e apresentação desses dados, bem como na transformação dos mesmos em informação, visto que os mais diversos setores como educação, negócios, saúde, indústria, tecnologia, entre outros, querem esse conhecimento oculto nos dados e não os dados como tal, por isso faz-se tão importante a aplicação da metodologia do KDD na descoberta de informações, lançando mão de algoritmos de *Data Mining* a fim de encontrar padrões nos dados.

As aplicações de *Big Data* têm por premissa básica a correlação e busca de dados em plataformas heterogêneas através de suas ferramentas de análise, para assim gerar resultados. Mascarenhas e Pilan (2016) destacam que a utilização de técnicas de análise do *Big Data* (*Big Data Analytics*) pode ser de suma importância para diversos campos, incluindo entre estes a educação, visto que a extração, a interpretação e a comparação dos dados obtidos podem auxiliar instituições de ensino a verificar as afinidades dos estudantes com determinadas disciplinas, qual a maior facilidade dos alunos, desenvolvimento do aluno no decorrer do curso, entre outros. Para mais, as análises de *Big Data* para a educação estão intrinsecamente ligadas à personalização do ensino, em especial no foco da individualidade de cada estudante, podendo assim adaptar o aprendizado de acordo com as necessidades de cada um, mas sem a separação do coletivo ao qual pertence.

Existem poucas aplicações de Mineração de Dados para a educação, que é um

campo que também gera cada vez mais dados, como dados dos professores, faculdade, horas de ensino, alunos, cursos, entre outros. Na educação, o *Data Mining* pode ser aplicado principalmente para prever o desempenho dos estudantes e para formular novas técnicas de aprendizagem, com o intuito de remover os pontos fracos do setor educacional e otimizar o desempenho do aluno. (GULATI; SHARMA, 2012)

No que concerne a Mineração de Dados Educacionais (EDM) no Brasil, pode-se afirmar que:

O Brasil tem uma grande oportunidade para promover a revolução da EDM e beneficiar milhares de alunos; em grande parte por causa do incentivo governamental ao uso da Educação-a-Distância (EaD). Através da coleta de dados em grande escala é possível criar modelos e fazer previsões que serão aplicáveis em qualquer ambiente virtual de aprendizagem e até mesmo em salas de aula convencionais. Para isso, é preciso que os dados das interações dos alunos com o material didático e com os professores e colegas nos ambientes de EaD sejam disponibilizados de forma padronizada e estruturada para a comunidade científica brasileira. Além disso, esses dados precisam incluir as informações necessárias para viabilizar a pesquisa e o estudo aprofundado da educação no país. Assim, acredita-se que a EDM tem grande potencial para ajudar o Brasil a se destacar no cenário educacional mundial através de ações que promovam o ensino eficaz nos ambientes de EaD e nas escolas através do uso de tecnologias educacionais que complementam o ensino em sala de aula.. (BAKER; ISOTANI; CARVALHO, 2011)

Na atualidade os estudantes estão cada vez mais imersos na Internet, o que por sua vez ocasiona na geração de muitos dados sobre eles, e através da Mineração de Dados e do *Big Data*, é possível extrair informações úteis desses dados que podem auxiliar a educação e formular estratégias apropriadas para garantir a boa aprendizagem do aluno e a qualidade na educação.

Posto isto, é fato que a análise de dados no campo da educação é uma realidade cada vez mais presente, trazendo elementos como aprendizagem adaptativa, onde o professor pode adaptar os materiais de ensino de acordo com as habilidades que o educando já possui; a melhoria no ensino; a constante autoavaliação do material ofertado aos alunos; registros da vida escolar do educando. Com efeito, a utilização de análise de dados pode ser considerada um grande benefício tanto para docentes como para alunos, em especial no tocante a compreensão de que cada mente humana é um mundo a parte com suas próprias necessidades. (HEKIMA, 2016)

Epilogando, a utilização de *Big Data* com a Mineração de Dados tem se tornado cada vez mais comum, e instituições de ensino estão aplicando a análise de aprendizado a fim de melhorar os serviços ofertados, notas dos estudantes, retenção, e possibilidade

de aproveitar os ciclos de *feedback* existentes entre professores e alunos. (PRAKASH; HANUMANTHAPPA; KAVITHA, 2014)

REFERÊNCIAS

BAKER, Ryan Shaun Joazeiro de.; ISOTANI, Seiji; CARVALHO, Adriana Maria Joazeiro Baker de. **Mineração de Dados Educacionais: Oportunidades para o Brasil**. Revista Brasileira de Informática na Educação, Volume 19, Número 2, 2011. Disponível em: < <http://br-ie.org/pub/index.php/rbie/article/view/1301>>. Acesso em: 31 mar. 2018.

BARBOSA, Anabela Aparecida Silva; CARVALHO, Rafael Nink de; ANDRADE, Fábio Santos. **Mineração de Dados em Ambientes Virtuais de Aprendizagem: Aportes para a Pesquisa em Educação a Distância**. Interfaces Científicas – Educação, Aracaju – v.6, n.1, p. 125-136, 2017. Disponível em: < <https://periodicos.set.edu.br/index.php/educacao/article/viewFile/4347/2421> >. Acesso em: 30 mar. 2018.

CÔRTEZ, Sérgio da Costa; PORCARO, Rosa Maria; LIFSCHITZ, Sérgio. **Mineração de Dados – Funcionalidades, Técnicas e Abordagens**. PUC-RioInfMCC10/02, 2002. Disponível em: < ftp://obaluae.inf.puc-rio.br/pub/docs/techreports/02_10_cortes.pdf>. Acesso em: 30 mar. 2018.

COSTA, Evandro; BAKER, Ryan S. J. d.; AMORIM, Lucas; MAGALHÃES, Jonathas; MARINHO, Tarsis. **Mineração de Dados Educacionais: Conceitos, Técnicas, Ferramentas e Aplicações**. Jornada de Atualização em Informática na Educação – JAIE 2012. Disponível em: <<http://www.br-ie.org/pub/index.php/pie/article/view/2341>>. Acesso em: 30 mar. 2018.

DIAS, Robson dos Santos. **Caracterização do Learning Analytics na Educação a Distância**. Seminário de Pesquisa e Inovação Tecnológica, 2017. Disponível em: < editora.iftm.edu.br/index.php/sepit/article/download/312/145>. Acesso em: 01 abr. 2018.

DREYER, Loiva. **Alfabetização: O Olhar de Paulo Freire**. X Congresso Nacional de Educação – EDUCERE – I Seminário Internacional de Representações Sociais, Subjetividade e Educação – SIRSSE, 2011. Disponível em: < http://educere.bruc.com.br/CD2011/pdf/5217_2780.pdf>. Acesso em: 09 jun. 2018.

ENOMURA, Bianca Yuki. Bem-vindo à era do Big Data. Superinteressante. 2014. Disponível em: < <https://repositorio.ufsc.br/bitstream/handle/123456789/133419/BIG%20DATA%20-%20Superinteressante.pdf?sequence=1> >. Acesso em: 30 mar. 2018.

GRINBERGAS, Daniella. **Como o Big Data pode ser usado na educação:** Como o uso de técnicas de análise de dados, escolas podem mapear o desempenho de seus alunos e melhorar a gestão administrativa, controlando problemas como a evasão. *Revista Educação*, edição 223, 2015. Disponível em: <<http://www.revistaeducacao.com.br/como-o-big-data-pode-ser-usado-na-educacao/>>. Acesso em: 01 abr. 2018.

GULATI, Pooja; SHARMA, Archana. **Educational Data Mining for Improving Educational Quality.** *IRACST – International Journal of Computer Science and Information Technology & Security*, vol. 2, n.3, 2012. Disponível em: <<https://ijcsits.org/papers/Vol2no32012/26vol2no3.pdf>>. Acesso em: 31 mar. 2018.

HAN, Jiawei; KAMBER, Micheline; PEI, Jian. **Data Mining: Concepts and Techniques.** 3rd ed. Morgan Kaufmann Publishers, 2012.

HEKIMA. Como Big Data Analytics ajuda na educação. *Big Data Business*, 2016. Disponível em: <<http://www.bigdatabusiness.com.br/como-o-big-data-analytics-ajuda-na-educacao-2/>>. Acesso em: 09 jun. 2018.

ISHWARAPPA, Kalbandi; ANURADHA, J. **A Brief Introduction on Big Data 5Vs Characteristics and Hadoop Technology.** *Procedia Computer Science* 48 (2015) 319 – 324. Disponível em: <https://ac.els-cdn.com/S1877050915006973/1-s2.0-S1877050915006973-main.pdf?_tid=8eb0d6ca-7ece-4d36-b76b-03cf10654ccf&acdnat=1522445220_657207b92e29654f47743e0cfee4d786>. Acesso em: 30 mar. 2018.

JINDAL, Rajni; BORAH, Malaya Dutta. **A Survey on Educational Data Mining and Research Trends.** *International Journal of Database Management Systems*, vol. 5, n.3, 2013. Disponível em: <<http://airccse.org/journal/ijdms/papers/5313ijdms04.pdf>>. Acesso em: 31 mar. 2018.

KRPAN, Divna; STANKOV, Slavomir. **Educational Data Mining for Grouping Students in E-learning System.** *Proceedings of the ITI 2012 34th Int. Conf. on Information Technology Interfaces*, 2012. Disponível em: <<http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6308006>>. Acesso em: 31 mar. 2018.

MARCONI, Marina de Andrade; LAKATOS, Eva Maria. **Fundamentos de Metodologia Científica.** 5^a Edição. São Paulo: Editora Atlas S.A., 2003.

MASCARENHAS, Leonardo N.; PILAN, José Rafael. **A Utilização do “Big Data Analytics” na Educação.** 5^a Jornada Científica e Tecnológica da FATEC de Botucatu, 2016. Disponível em: <<http://www.fatecbt.edu.br/ocs/index.php/VJTC/VJTC/paper/viewFile/734/914>>. Acesso em: 01 abr. 2018.

MERKER, Júlia. Seis cursos de Big Data em faculdades. Baguete, 2014. Disponível em: < <https://www.baguete.com.br/noticias/11/07/2014/seis-cursos-de-big-data-em-faculdades>>. Acesso em: 01 abr. 2018.

MOUTINHO, Sofia. Quanta informação há no mundo? Ciência Hoje. 2011. Disponível em: < http://www.cienciahoje.org.br/noticia/v/ler/id/1487/n/quanta_informacao_ha_no_mundo>. Acesso em: 30 mar. 2018.

NITHYA, P.; UMAMAHESWARI, A.; UMADEVI, A. **A Survey on Educational Data Mining in Field of Education**. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Vol. 5, 2016. Disponível em: < <http://ijarcet.org/wp-content/uploads/IJARCET-VOL-5-ISSUE-1-69-78.pdf>>. Acesso em: 31 mar. 2018.

OLIVEIRA, Jayr Figueiredo de. **T.I.C.: Tecnologias da Informação e da Comunicação**. São Paulo: Érica, 2003.

PATRICIO, Thiago Seti. **Ferramenta Para Mineração de Regra de Associação Web**. 2014. 75f. Trabalho de Conclusão de Curso (Tecnologia em Banco de Dados) – Faculdade de Tecnologia de Lins Prof.º Antonio Seabra – Lins – SP. Disponível em: < <http://fateclins.edu.br/v4.0/informacoesTrabalhoGraduacaoAluno.php?idTG=197>>. Acesso em: 30 mar. 2018.

PAVAN, Fabiana Cruz. **Diversidade Cultural e Aprendizagem Significativa na Construção da Cidadania**. Revista Científica Multidisciplinar Núcleo do Conhecimento. Ano 1. Vol. 9, p. 632-647, 2016. Disponível em: <<https://www.nucleodoconhecimento.com.br/educacao/diversidade-cultural-e-aprendizagem-significativa-na-construcao-da-cidadania>>. Acesso em: 09 jun. 2018.

PRAKASH, B. R.; HANUMANTHAPPA, M.; KAVITHA, Vasantha. **Big Data in Educational Data Mining and Learning Analytics**. International Journal of Innovative Research in Computer and Communication Engineering, vol. 2, Issue 12, 2014. Disponível em: <<http://www.rioi.com/open-access/big-data-in-educational-data-mining-andlearning-analytics.pdf>>. Acesso em: 01 abr. 2018.

PRODANOV, Cleber Cristiano; FREITAS, Ernani Cesar. **Metodologia do Trabalho Científico: Métodos e Técnicas da Pesquisa e do Trabalho Acadêmico**. 2ª edição. Novo Hamburgo – Rio Grande do Sul: Universidade Feevale, 2013.

RANGRA, Kalpana; BANSAL, K. L. **Comparative Study of Data Mining Tools**. International Journal of Advanced Research in Computer Science and

Software Engineering. 2014. Disponível em: < <https://pdfs.semanticscholar.org/dd3c/89280a078131cd2bc1be6c3c6db2bc38c58f.pdf>> Acesso em: 30 mar. 2018.

SAAS, Amjad Abu. **Educational Data Mining & Students's Performance Prediction**. International Journal of Advanced Computer Science and Applications, vol. 7, n.7, 2016. Disponível em: < https://thesai.org/Downloads/Volume7No5/Paper_31-Educational_Data_Mining_Students_Performance_Prediction.pdf >. Acesso em: 31 mar. 2018.

SABBATINI, Renato M. E. **Ambiente de Ensino de Aprendizagem via Internet – A Plataforma Moodle**. Instituto EduMed, 2007. Disponível em: <<http://www.ead.edumed.org.br/file.php/1/PlataformaMoodle.pdf>>. Acesso em: 09 jun. 2018.

SCAICO, Pasqueline Dantas; QUEIROZ, Ruy José G. B. de; SCAICO, Alexandre. **O conceito *big data* na educação**. 3º Congresso Brasileiro de Informática na Educação (CBIE 2014) – 20º Workshop de Informática na Escola (WIE 2014). Disponível em: < <http://www.br-ie.org/pub/index.php/wie/article/view/3115>>. Acesso em: 30 mar. 2018.

SIN, Katrina; MUTHU, Loganathan. **Application of Big Data in Education Data Mining and Learning Analytics – A Literature Review**. ICTACT Journal on Soft Computing: Special Issue on Soft Computing Models for Big Data, vol. 5, 2015. Disponível em: < http://ictactjournals.in/paper/IJSC_V5_I4_paper6_1035_1049.pdf>. Acesso em: 01 abr. 2018.

SRIVASTAVA, Jaya; SRIVASTAVA, Abhay Kumar. **Data mining in education sector: a review**. International Journal of Advanced Networking Applications, Special Conference Issue, National Conference on Current Research Trends in Cloud Computing & Big Data, p. 184–190, 2013. Disponível em: < <https://pdfs.semanticscholar.org/d91e/ee207844e8d3a2dfc7308fa243bbc1a8a3ae.pdf>>. Acesso em: 31 mar. 2018.