

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

Big Data nas Escolas
O processo de Ensino Adaptativo

Fernanda de Jesus Amorim

JUIZ DE FORA
FEVEREIRO, 2016

Big Data nas Escolas

O processo de Ensino Adaptativo

FERNANDA DE JESUS AMORIM

Universidade Federal de Juiz de Fora
Instituto de Ciências Exatas
Departamento de Ciência da Computação
Bacharelado em Ciência da Computação

Orientador: Liamara Scortegagna

JUIZ DE FORA
FEVEREIRO, 2016

BIG DATA NAS ESCOLAS
O processo de Ensino Adaptativo

Fernanda de Jesus Amorim

MONOGRAFIA SUBMETIDA AO CORPO DOCENTE DO INSTITUTO DE CIÊNCIAS EXATAS DA UNIVERSIDADE FEDERAL DE JUIZ DE FORA, COMO PARTE INTEGRANTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE BACHAREL EM CIÊNCIA DA COMPUTAÇÃO.

Aprovada por:

Liamara Scortegagna
Doutora em Engenharia de Produção (UFSC)

Victor Stroele de Andrade Menezes
Doutor em Engenharia de Sistemas e Computação (UFRJ)

Luciana Conceição Dias Campos
Doutora em Engenharia Elétrica (PUC - Rio)

JUIZ DE FORA
<<DIA>> DE FEVEREIRO, 2016

*Primeiramente, a Deus por ter me dado saúde,
e força para superar as dificuldades.*

À minha avó, pelo apoio e sustento.

*Aos meus pais, irmãos e amigos pelo apoio e
amor incondicional.*

*Ao meu namorado pela compreensão e cuida-
dos.*

*A esta Universidade pela oportunidade de fa-
zer o curso.*

*À professora Liamara pelo suporte, pelas suas
correções e incentivos.*

*A todos que direta ou indiretamente fizeram
parte da minha formação, o meu muito obri-
gada.*

Resumo

As TICs (Tecnologias de Informação e Comunicação) vêm sendo amplamente utilizadas, inclusive na Educação, o que gera uma grande quantidade de dados sobre o processo de ensino-aprendizagem. Para processar e analisar todo esse volume de informações estão sendo utilizadas ferramentas de Big Data.

A fim de dar compreensão sobre seu conceito e sua funcionalidade, este trabalho apresenta as definições mais relevantes do termo Big Data, bem como as dimensões e variáveis que o fundamentam: os 5Vs (volume, variedade, velocidade, veracidade e valor), e cita exemplos de sua utilização nas mais variadas áreas de atuação da sociedade, com ênfase em como essa tecnologia tem sido aplicada à Educação. Isso para ilustrar como ocorre o processo de Ensino Adaptativo para aprimorar os modelos educacionais.

Algumas ferramentas de Ensino Adaptativo são apresentadas para exemplificar a aplicação de Big Data na Educação. Isso para mostrar seus benefícios, usos atuais e desafios em sua implementação no contexto da Educação.

O objetivo deste trabalho é divulgar essa tendência para que seja possível iniciar futuras discussões sobre o uso das técnicas de análise de Big Data como apoio à melhoria do processo de ensino-aprendizagem no Brasil.

Palavras-chave: Big Data, Ensino Adaptativo, Analíticos, Educação, Aprendizagem, Predição.

Abstract

ICT (Information and Communication Technologies) have been widely used, including in education, which generates a lot of data on the teaching-learning process. To process and analyze all this volume of information is being used Big Data tools.

In order to give understanding of its concept and functionality, this work presents the most relevant definitions of the term Big Data, as well as the dimensions and variables that are based: the 5Vs (volume, variety, velocity, veracity and value), and quotes examples of their use in various areas of activity of the company, with an emphasis on how this technology has been applied to education. This is to illustrate how the process of adaptive learning to improve educational models.

Some Adaptive Teaching tools are presented to exemplify the application of Big Data in Education. This to show its benefits, current uses and challenges in its implementation in the context of education.

The objective of this work is to promote this trend so that you can start further discussions on the use of Big Data analysis techniques to support the improvement of the teaching-learning process in Brazil.

Keywords: Big Data, Adaptative learning, Analytics, Education, Learning, Prediction.

Agradecimentos

Primeiramente agradeço a Deus, o centro e o fundamento de tudo em minha vida, por renovar a cada momento a minha força e disposição e pelo discernimento concedido ao longo dessa jornada.

Aos meus pais, aos meus irmãos, a minha avó Teresinha e a toda minha família que, com muito carinho e apoio, não mediram esforços para que eu chegasse até esta etapa da minha vida.

Ao meu namorado Matheus Oliveira pelo carinho, pela compreensão e, principalmente, pelos cuidados durante esse período do TCC.

A todos os meus colegas do curso de Ciência da Computação por partilharem da experiência desafiante da vida acadêmica comigo. Peço a Deus que os abençoe grandemente, preenchendo seus caminhos com muita paz, amor, saúde e prosperidade.

Aos docentes do curso de Ciência da Computação pela convivência harmoniosa, pelas trocas de conhecimento e experiências que foram tão importantes na minha vida acadêmica/pessoal e contribuíram para o meu novo olhar profissional.

À minha orientadora, pelo empenho dedicado à elaboração deste trabalho, pelas suas correções e incentivos.

A todos que direta ou indiretamente fizeram parte da minha formação, o meu mais sincero agradecimento.

*Já cheguei até aqui
Eu não posso desistir
Devo achar forças pra andar
As lutas vieram, mas também passaram
Vendavais tentaram me derrubar.
Mas busquei a Cristo, meu fiel amigo
Que esteve comigo todo dia
Mesmo que chorando, mesmo que sofrendo
Nada me faltou, tudo Deus proveu
E me deu vitórias ontem, hoje e me dará
no amanhã*

*(Leandro Oliveira e Janne Gonçalves -
2002)*

Sumário

Lista de Abreviações	7
1 Introdução	8
1.1 Justificativa	12
1.2 Objetivos	13
1.3 Metodologia	13
1.4 Estrutura do trabalho	14
2 Big Data	15
2.1 Conceito e histórico	15
2.1.1 O que são dados	15
2.1.2 Evolução dos dados	17
2.1.3 O que é Big Data	26
2.2 Aplicações	34
2.2.1 Recursos Humanos	35
2.2.2 Propaganda e Marketing	36
2.2.3 Localização	37
2.2.4 Esporte	38
2.2.5 Segurança	39
2.2.6 Economia	42
2.2.7 Dados climáticos	43
2.2.8 Comercial	44
2.2.9 Saúde	44
2.2.10 Turismo	46
3 Big Data nas Escolas	47
3.1 Conceito e prática do Ensino Adaptativo	48
3.2 Aplicação de Big Data no Ensino	57
3.2.1 Knewton	58
3.2.2 Khan Academy	58
3.2.3 Geekie Lab	60
3.2.4 Adaptativa	61
3.2.5 Enem Quiz	62
3.2.6 Ismart Online	63
3.2.7 Moodle	64
4 Análises	65
5 Considerações finais	73
Referências Bibliográficas	75

Lista de Abreviações

ACID	Atomicity, Consistency, Isolation e Durability	Atomicidade
AVA	Ambiente Virtual de Aprendizagem	
BASE	Basically Available, Soft state, Eventually consistency	
BI	Business Intelligence	
CAT	Common Admission Test	
CIO	Chief Information Officer	
DAS	Domain Awareness System	
EAD	Educação a Distância	
ERP	Enterprise Resource Planning	
HTML	HyperText Markup Language	
IA	Inteligência Artificial	
IDC	Internet Data Corporation	
IETF	Internet Engineering Task Force	
ios	iphone Operation System	
IP	Internet Protocol	
MIT	Massachusetts Institute of Technology	
Moodle	Modular Object Oriented Distance Learning	
MRP	Material Requirements Planning	
NoSQL	Not Only SQL	
PC	Personal Computer	
PDA	Personal Digital Assistant	
SGDB	Sistema de Gerenciamento de Banco de Dados	
SI	Sistema de Informação	
SQL	Structured Query Language	
TIC	Tecnologia de Informação e Comunicação	
TRI	Teoria de Resposta ao Item	
XML	eXtensible Markup Language	

1 Introdução

A produção e a troca de informações entre pessoas, empresas e governos ocorrem em quantidade e intensidade crescentes e, por esse motivo, o mundo moderno está sendo denominado de Sociedade da Informação (SocInfo), termo cunhado pelo sociólogo Manuel Castells (2003). Os avanços tecnológicos acrescentam a esse quadro a automação de processos que exigem dos trabalhadores cada vez mais qualificação e conhecimento.

A área de Tecnologia da Informação (TI) através de empresas, centros de pesquisas e universidades têm uma contribuição expressiva para os mais variados segmentos da sociedade, incluindo a Educação, colaborando com a educação inicial, inclusão digital e o empreendedorismo, entendido como uma das variantes da qualificação.

O método de ensino tradicional, em que a lousa, o giz e o material didático a serviço do professor, cujo trabalho limita-se a repetir conceitos, fórmulas, esquemas e modelos prontos e acabados, não deixa espaço para a criação por parte dos alunos e, por isso, aos poucos está sendo substituído por aulas mais interativas e dinâmicas. Isso porque as instituições de ensino têm adotado o uso da Tecnologia da Informação (TI) dentro e fora da sala de aula, o que torna a aula mais motivadora, além de conduzir o aluno a inúmeras descobertas.

A Tecnologia da Informação (TI) pode ser definida como o conjunto de atividades e soluções providas por recursos de computação que visam a permitir o armazenamento, o acesso e o uso das informações para auxiliar a tomada de decisão. A TI aplicada à Educação, inclusive na formação, treinamento e reciclagem, deve ser capaz de garantir a professores e alunos as ferramentas necessárias para melhorar a qualidade do ensino. Nesse contexto, o professor busca se transformar em um mediador entre os alunos e o processo de construção do conhecimento, e os estudantes podem encontrar estímulos para que o interesse pelo aprendizado esteja mais presente em seu cotidiano e, desse modo, colaborar com outras pessoas, trocar todos os tipos de informações e, ainda, aprender a construir o próprio conhecimento, em um processo cumulativo de ajuda mútua e de percepção compartilhada de problemas e necessidades.

Inicialmente, entendia-se por Tecnologia na Educação apenas computadores ligados em redes locais e remotas, que serviam de repositório para os poucos materiais digitalizados da instituição, para o uso de *softwares* educacionais e para registrar frequência e notas dos alunos, por exemplo. Em seguida, veio a *Internet*, que revolucionou a forma como os conteúdos poderiam ser apresentados aos alunos em sala de aula e fora dela. Entre as várias mudanças pode-se destacar: buscas em *sites* de pesquisas, como por exemplo o *Google*, *sites* de outras instituições de ensino, vídeos que podem complementar e ilustrar as explicações dadas em salas de aula, fóruns, *chats*, *blogs* educacionais, exercícios *online*, entre outros.

Com o apoio das Tecnologias de Informação, a escola sai do ambiente tradicional e passa a funcionar em qualquer lugar onde haja acesso à *Internet*. Os dispositivos móveis como *tablets*, celulares e *notebooks* destacam-se como importantes ferramentas de ensino, principalmente no que diz respeito à interatividade, ao acesso a conteúdo e à economia de papel, contribuindo para um mundo mais sustentável. Na Sociedade do Conhecimento é muito difícil manter um livro didático impresso atualizado, mas com a disponibilidade de acesso à *Internet* isso é possível.

Devido à utilização da tecnologia no segmento da Educação, ocorreu uma evolução da modalidade de ensino não presencial. A Educação a Distância (EAD) tornou-se uma ferramenta estratégica, especialmente em um país de dimensões geográficas extensas como o Brasil, onde a formação e a geração do conhecimento ainda estão limitadas aos grandes centros urbanos. As soluções de *e-learning* ou educação *online* ajudam a democratizar o acesso ao ensino, promovendo a inclusão social, permitindo que crianças, jovens e adultos, em suas respectivas comunidades, tenham acesso às mesmas oportunidades.

Além disso, as ferramentas tecnológicas disponíveis para esta modalidade de ensino possibilitam acompanhar como está ocorrendo a aprendizagem individual ou de um grupo de alunos, verificar o desempenho, a interação, a forma de acesso, o tipo de dispositivo ou a tecnologia que os alunos estão utilizando e, ainda, por onde estes alunos “andam” virtualmente no espaço de ensino e aprendizagem ou mesmo na *Internet*.

Todas as atividades ou ações desempenhadas por um aluno e descritas acima geram informações e sabemos que estas possuem valores inestimáveis quando tratadas e in-

interpretadas de forma correta e com objetivos bem definidos. Trabalhar essas informações, utilizando-as adequadamente, é tarefa da Tecnologia da Informação (TI).

Segundo a pesquisa *NMC Horizon Report: 2014 Higher Education Edition* sobre tendências na educação mundial, publicada no início de 2014, o uso de dados e informações com objetivo de melhorar o processo de ensino e aprendizagem ou ainda para a criação de aulas personalizadas se tornará mais comum até o ano de 2020. Isso só é possível porque cada vez mais os estudantes têm feito tarefas *online* e isso deixa um rastro de dados e estatísticas na *Web*. Essas informações podem ser coletadas e analisadas a fim de criar técnicas de Ensino Adaptativo e, assim, melhorar o aprendizado.

Esse tipo de análise é chamado de Big Data¹, termo utilizado para descrever grandes volumes de dados que utilizam ferramentas especiais para encontrar e analisar em tempo hábil o que de fato é útil dentre tudo o que foi coletado e armazenado (Alecrim, 2015), e já é utilizado em universidades como a de Connecticut nos Estados Unidos.

Conforme apontou a pesquisa *NMC Horizon Report: 2014 Higher Education Edition*, citada anteriormente, o Big Data deve se expandir cada vez mais pelo mundo nos próximos cinco anos. Aqui no Brasil, no entanto, o seu uso ainda é remoto e feito de maneira precária, sem aproveitar o máximo que a técnica oferece.

Como consequência do tratamento das informações coletadas, tem-se o Ensino Adaptativo, que consiste em métodos de aprendizagem que utilizam computadores como dispositivos de ensino interativo que se adaptam de acordo com as necessidades específicas de cada aluno, mediante suas respostas a perguntas, tarefas e experiências. A adaptação é dinâmica e ocorre a partir das competências e das habilidades do aluno. Algoritmos sugerem as respostas para algumas decisões sobre qual conteúdo recomendar a um aluno, em qual sequência e para qual objetivo. Dessa forma, os professores não precisam tomar todas as decisões, fazendo isso apenas para os casos mais críticos e ajustando as sugestões que julgarem equivocadas. Cada professor consegue ser um tutor particular para toda uma sala de aula e o sistema de ensino personalizado se torna mais inclusivo.

Os sistemas de Ensino Adaptativo podem ser implementados na *Internet* para uso em aplicações de Educação a Distância e de colaboração em grupo. Somente agora,

¹Optou-se pela utilização do termo em inglês

o campo de EAD está incorporando aspectos da aprendizagem adaptativa. Antes disso, os sistemas eram capazes de fornecer *feedback* automatizado para os estudantes, para os quais eram apresentadas questões a partir de um banco de perguntas pré-selecionado. Essa abordagem, no entanto, não tem a orientação que os professores em sala de aula podem fornecer. As tendências atuais do EAD sinalizam para o uso de Ensino Adaptativo para implementar um comportamento dinâmico e inteligente ao ambiente de aprendizagem, de forma a oferecer a cada aluno a melhor experiência e o melhor aproveitamento da plataforma.

Uma extensão da EAD, a colaboração em grupo surge como uma nova área de pesquisa de aprendizagem adaptativa. Através dela é possível realizar o agrupamento automático de usuários com os mesmos interesses e personalizar *links* para fontes de informação, baseados em interesses declarados do usuário ou hábitos de navegação do mesmo.

O Ensino Adaptativo também pode ser uma proposta complementar às aulas presenciais, em que são trabalhados debates e outras ações coletivas, com percursos individuais de aprendizagem que vão sendo elaborados por meio das escolhas do aluno, de um *software* que acompanha o seu desempenho e, quando possível, de tutores. Essa concepção teve início em 1970, junto com o movimento de Inteligência Artificial (IA) (Mayer-Schonberger e Cukier, 2015). Nessa época, acreditava-se que os computadores seriam capazes de adaptar-se, assim como os humanos, e com isso seria possível cumprir a premissa básica da aprendizagem adaptativa: o sistema ser capaz de se ajustar ao método de aprendizagem do aluno, resultando em uma experiência mais eficaz de aprendizado para ele (Mayer-Schonberger e Cukier, 2015). As principais barreiras para a popularização dessas ferramentas na década de 70 foram o custo, o tamanho dos computadores e a interface de usuário, que não era propícia para o processo de aprendizagem.

A evolução da tecnologia tem seguido a Lei de Moore (Ulisses, 2015), que prevê que a capacidade de processamento e armazenamento dobra a cada 18 meses, portanto a capacidade de armazenamento tende a ser cada vez maior. Além disso, quando se aplica a Lei de Parkinson (Ulisses, 2015) à computação, determina-se que toda a capacidade disponível tende a ser ocupada, e isso significa mais dados para armazenar e, principal-

mente, analisar, o que tem se tornado inviável com ferramentas convencionais. Com isso, surge a importância das ferramentas de análise de Big Data, que permitem o tratamento de grandes quantidades de informações, estruturadas ou não, para gerar novas análises e entendimento sobre as informações existentes e apresentar uma nova forma de planejar e fazer Educação através do Ensino Adaptativo.

1.1 Justificativa

Para se manter integrada à realidade que a cerca, a escola tem que se adaptar aos recursos e ferramentas de informática e tem que evoluir para integrá-los na ação educativa normal e, com isso, facultar aos alunos uma familiarização no acesso a repositórios de dados e às múltiplas oportunidades de interação social. Para isso, ela tem que melhorar o processo de ensino e aprendizagem, acompanhando a evolução da utilização das Tecnologias de Informação na Educação e transformando o ensino tradicional em Ensino Adaptativo.

Por isso, é preciso conhecer primeiramente as tendências tecnológicas que estão surgindo nesse cenário. O Big Data é uma delas e a grande importância pedagógica da sua utilização nas escolas é que os alunos podem aprender fazendo coisas, em vez de aprender ouvindo dizer como as coisas devem ser feitas, ou seja, eles participam diretamente da construção do seu conhecimento a partir de um aprendizado personalizado de acordo com seu rendimento e suas habilidades.

O Ensino Adaptativo representa a utilização do Big Data nas escolas e busca aperfeiçoar o processo de aprendizagem, auxiliando os alunos a construírem seus próprios percursos de aprendizagem, através de recursos tecnológicos para educação *online*. O objetivo do Ensino Adaptativo é instruir cada aluno com base em suas necessidades individuais e complementar as aulas presenciais a partir de avaliações contínuas e individuais.

A proposta de aprofundar o conhecimento sobre Big Data e Ensino Adaptativo apresentados neste estudo se justifica pela importância acima apresentada.

1.2 Objetivos

Com este trabalho, busca-se pesquisar e descrever a utilização do Big Data nas escolas, apresentar como ocorre o processo de Ensino Adaptativo, a fim de divulgá-lo e ilustrar o potencial do Big Data para melhorar o cenário educacional e os desafios que estão sendo enfrentados em sua implementação no contexto da Educação.

Como objetivos específicos, podem ser citados:

- a) Realizar uma pesquisa sistemática sobre Big Data e os assuntos tratados por essa tecnologia;
- b) Estudar e apresentar soluções de Big Data relacionadas à Educação propostas na literatura;
- c) Discorrer sobre as ferramentas disponíveis e apresentar como elas podem aprimorar o ensino a partir da metodologia do Ensino Adaptativo.

1.3 Metodologia

Com o intuito de alcançar os objetivos descritos, será realizada uma pesquisa exploratória sobre os temas Big Data e Ensino Adaptativo, a fim de estabelecer a base teórica necessária para a análise e apresentação de ferramentas de Big Data voltadas para Educação. Com isso, busca-se auxiliar as escolas na escolha de uma dessas ferramentas para a inclusão da metodologia de Ensino Adaptativo em suas instituições.

Esse tipo de pesquisa, segundo Gil (2008), busca trazer mais informações sobre um tema através de pesquisas bibliográficas. Em geral, esse tipo de pesquisa envolve:

- a) Um levantamento bibliográfico, que utiliza artigos, monografias, dissertações, teses, livros, periódicos e pesquisas na *Internet* como fontes de pesquisa;
- b) A análise de exemplos para auxiliar no entendimento do tema.

Através de um levantamento bibliográfico serão identificadas as principais características do Big Data para dar compreensão sobre seu conceito e funcionamento. Também

será mostrado como o Big Data está sendo utilizado na Educação para transformar o ensino tradicional em Ensino Adaptativo, e serão enumeradas as principais ferramentas adaptativas disponíveis.

1.4 Estrutura do trabalho

O trabalho está estruturado em quatro capítulos, como a seguir:

No primeiro capítulo, de introdução, é realizada a contextualização do tema e apresentada a justificativa, bem como os objetivos estabelecidos. Ainda nesta seção há a metodologia utilizada na pesquisa e a maneira como o trabalho está estruturado.

Os principais conceitos acerca do termo Big Data e seu histórico estão descritos no segundo capítulo, bem como as características nas quais ele é baseado. Nesta seção, também são listadas as definições mais relevantes e apresentados os principais exemplos de utilização das técnicas de Big Data nas mais variadas áreas de atuação a fim de dar compreensão sobre seu conceito e sua funcionalidade.

No terceiro capítulo, é abordado como o Big Data está sendo utilizado nas escolas para aprimorar os modelos educacionais, citando as técnicas e metodologias adotadas para garantir um melhor aprendizado. Além disso, são apresentadas algumas ferramentas de Ensino Adaptativo para exemplificar a aplicação de Big Data na Educação.

O quarto capítulo é destinado à análise do estudo, em que é realizada uma sucinta apresentação da pesquisa.

Por fim, o quinto capítulo apresenta as considerações finais a respeito do tema abordado neste trabalho e algumas propostas para trabalhos futuros.

2 Big Data

Segundo o estudo Universo Digital, realizado pela EMC (uma das principais empresas de soluções de Big Data do mundo), entre 2013 e 2020, o volume de informações digitais passará de 4,4 trilhões de *gigabytes* para 44 trilhões. Isso se tornou possível porque o custo do armazenamento de dados caiu exponencialmente nos últimos anos.

Devido à falta de tecnologias capazes de processar grandes quantidades de informações, em 2013, apenas 5% dos dados considerados úteis foram analisados. Esse panorama mudou quando as soluções de Big Data se tornaram mais acessíveis, permitindo a análise de dados digitais em volume, variedade e velocidade inéditos, e facilitando a identificação do que realmente é útil entre todos os dados coletados. Na prática, a tecnologia permite analisar qualquer tipo de informação digital, em tempo real, sendo fundamental para a tomada de decisões.

Como não há uma definição única para o termo Big Data, neste capítulo, são apresentados os conceitos acerca do termo, bem como as características que o fundamentam. Além disso, são listadas as definições mais relevantes e alguns exemplos de aplicações do Big Data a fim de dar compreensão sobre seu conceito e sua funcionalidade.

2.1 Conceito e histórico

Em tradução livre, Big Data significa “Grandes Dados” e apesar de não haver um consenso entre os especialistas a respeito do conceito desse termo, sabe-se que ele representa a capacidade de processar um volume inédito de dados digitais variados e analisá-los em tempo hábil para tomada de decisões. Para entender o que é Big Data, é preciso compreender o que são dados e sua evolução ao longo dos anos.

2.1.1 O que são dados

Em uma definição generalizada do termo, dados são um conjunto de informações que podem ou não estar organizadas. Informações essas que podem ser quantitativas, qualita-

tivas, categóricas ou indefinidas. Isso depende das características da ciência que as define. Em Ciência da Computação, dado é a menor informação fornecida ou processada por um computador.

Setzer (2015) define dado como uma sequência de símbolos quantificados ou quantificáveis. Para o autor, quantificável significa que algo pode ser quantificado e depois reproduzido sem que se perceba a diferença para com o original. Por exemplo, um texto é um dado e as letras são símbolos quantificados. Fotos, figuras, sons e vídeos também são considerados dados, pois podem ser quantificados ao serem introduzidos em um computador.

Os dados podem ser descritos através de representações formais e estruturais. Sendo quantificáveis, eles podem ser armazenados em um computador e processados por ele. O processamento de dados em um computador limita-se exclusivamente a manipulações estruturais dos mesmos, e é feito por meio de programas. No caso de textos, essas manipulações podem ser exemplificadas por formatação, ordenação, comparação com outros textos, estatísticas de palavras empregadas etc (Setzer, 2015).

Já os autores Davenport et al (1998) descrevem dado como “um conjunto de fatos discretos, objetivos, sobre eventos” e estabelecem que “os dados por si mesmos têm pouca intenção ou relevância”, pois “[...] não existe qualquer significado inerente aos dados. Os dados descrevem somente uma parte do que aconteceu”. Isso porque os dados são apenas representações simbólicas e não possuem qualquer conexão com o que descrevem. Eles podem representar informações ou podem não representar nada, sendo assim, impossível extrair deles qualquer tipo de informação. Por exemplo, uma tabela de cidades e temperaturas em chinês não tem nenhum significado para quem não lê ou não compreende esse idioma (Setzer, 2015).

Na terminologia adotada pela ciência que trata do processamento automático da informação, dado é o “elemento de informação, ou representação de fatos ou instruções, em forma apropriada para armazenamento, processamento ou transmissão por meios automáticos” (Ferreira et al, 1999).

Para Frangomeni (1996), dados são “elementos básicos, passíveis de serem expressos como uma determinada combinação de sinais que têm significado através de um

código, e que, estruturados, podem conter informação”.

Correia (2009) descreve o dado como uma unidade elementar da informação:

“Uma unidade manifestante de um valor semântico básico que possui, portanto, pouca significação quando tomada de forma isolada. Porém, quando um dado é combinado a outros dados seus valores interagem e dão origem a significações mais complexas. As características fundamentais dos dados seriam, então a transportabilidade e a capacidade de agrupamento de seus valores semânticos elementares. ”

“Dados são importantes para as organizações em grande parte, naturalmente, porque constituem matéria prima essencial para a criação de informação” (Davenport et al, 1998). Informação é todo o conjunto de dados devidamente ordenados e organizados de forma a terem significado, ou seja, é o conhecimento produzido como resultado do processamento de dados.

Ferreira et al (1999) definem informação como o “ato de informar (-se), ou seja, o ato de tomar conhecimento, de inteirar-se ou instruir-se sobre algo”. Para os autores, informação significa “dados acerca de alguém ou algo” ou “conhecimento, participação”. Ainda de acordo com Ferreira et al (1999), a informação é “a medida da redução de incerteza, sobre determinado estado de coisas, por intermédio de uma mensagem”.

Correia (2009) identifica uma condição necessária entre os conceitos de dado e informação: “[...] não há informação sem dados, e dados não tem significância real antes de se tornarem informação. ”

2.1.2 Evolução dos dados

Marr (2015) afirma que a história do armazenamento de dados começou na História Antiga. Segundo o autor, o primeiro exemplo comprovado de armazenamento de dados ocorreu na Era Paleolítica em 1800 a.C., quando as tribos marcavam com palitos ou ossos as atividades comerciais e os suprimentos, e os comparavam para realizarem cálculos rudimentares e, com isso, determinarem quanto tempo duraria o suprimento de comida.

O ábaco foi o primeiro dispositivo construído especificamente para a realização de cálculos em 2400 a.C. Nesse período, também ocorreu a construção das primeiras bibliotecas, que continham pergaminhos que coletavam dados do Mundo Antigo e foram as primeiras tentativas de armazenamento de dados em massa de que se tem relatos (Marr,

2015).

A seguir, é descrito o histórico desenvolvido por Press (2013) para a *Forbes* (considerada a revista de negócios e economia mais conceituada do mundo) com os principais acontecimentos relacionados à evolução dos dados e seu armazenamento. Através desse histórico, este trabalho busca apresentar a evolução do pensamento e da inovação que levaram ao surgimento do Big Data, bem como as bases a partir das quais o termo foi fundado.

O computador mecânico mais antigo descoberto é o *Antikythera Mechanism*, construído pelos gregos para fins astrológicos e acompanhamento do ciclo dos Jogos Olímpicos no período de 100 d.C. a 200 d.C (Marr, 2015).

Já no contexto de análise de dados, John Graunt² publicou o livro *Natural and Political Observations upon the Bills of Mortality* em 1662, onde a partir de dados sobre enterros em Londres, o cientista criou uma teoria para projetar um sistema de alerta precoce para a peste bubônica na Europa. Essa foi a primeira experiência de análise estatística de dados.

Em 1865, Richard Millar Devens realizou o primeiro estudo sobre a análise de dados de negócios para fins comerciais. Nele, o cientista utilizou o termo *Business Intelligence* (BI) pela primeira vez e descreveu como o banqueiro Henry Fumace conseguiu vantagens sobre os concorrentes através da coleta e análise de informações estruturadas relevantes para suas atividades comerciais (Marr, 2015).

Em 1880, o *Instituto US Census Bureau* enfrentou um problema para processar os dados recolhidos no censo desse ano. Estimava-se que todo o processo demoraria cerca de 8 anos. A expectativa para o censo de 1890 era ainda mais pessimista: 10 anos para processar todos os dados, o que significa que as estatísticas já estariam ultrapassadas, uma vez que em 1900 outro censo seria realizado. Diante disso, o engenheiro Herman Hollerith³ criou a *Hollerith Tabulating Machine*, uma máquina que realizava as operações de recenseamento da população através da leitura de cartões perfurados. Isso reduziu o trabalho que antes era estimado em 10 anos para 3 meses. Ao criar os cartões perfurados que serviam de entrada de informação para os computadores da época, Hollerith se tornou

²John Graunt. Fonte: https://pt.wikipedia.org/wiki/John_Graunt

³Herman Hollerith. Fonte: <http://www.columbia.edu/cu/computinghistory/hollerith.html>

o precursor da computação moderna automatizada e foi um dos fundadores da IBM.

Em 1944, Fremon Rider⁴ publicou uma das primeiras tentativas de mensurar a quantidade de informação que estava sendo produzida. Ele observou que para armazenar tudo, as bibliotecas americanas teriam que dobrar sua capacidade a cada 16 anos. Isso o levou a especular que em 2040, a Biblioteca da Universidade de Yale teria 200 milhões de livros. Isso exigiria cerca de seis mil pessoas para catalogar todo o acervo da biblioteca.

Em 1956, Fritz- Rudolf Güntsch, um físico alemão, desenvolveu o conceito de Memória Virtual, que permitiu o processamento de dados sem restrição de memória ao gerenciar o armazenamento para esconder os detalhes do usuário (Press, 2013).

Apesar de já ter sido citado em 1865 por Richard Millar Devens, o termo *Business Intelligence* (BI) foi definido em 1958 pela IBM como “a capacidade de apreender as inter-relações dos fatos apresentados, de forma a orientar a ação para um objetivo desejado”.

Em 1961, devido à grande quantidade de pesquisas científicas criadas, Derek Price transformou todas elas em periódicos, numa tentativa de gerenciar a crescente base de conhecimento que estava sendo gerada. A isso ele deu o nome de Revolução Científica (Press, 2013).

A IBM também foi pioneira no reconhecimento de fala. Em 1962, ela apresentou a *Shoebbox Machine*⁵, capaz de interpretar números (de até dez dígitos) e dezesseis palavras faladas (no idioma Inglês) em informação digital.

Em 1963, ainda não havia sido encontrada uma solução organizacional para o problema encontrado por Price dois anos antes: armazenamento organizado do conhecimento científico para obter informações. Esse problema não era exclusividade do ramo científico, o mesmo estava acontecendo também no setor empresarial. Por isso, em 1966, as organizações começaram a projetar, desenvolver e implementar sistemas de computação centralizada, o que permitiu automatizar seus sistemas de inventário (Press, 2013).

Em 1970, Edgar F. Codd publicou um artigo, em que mostrou como informações armazenadas em grandes bancos de dados poderiam ser acessadas sem saber como a informação foi estruturada ou onde ela estava localizada no banco de dados. Tratava-se do

⁴Fremon Rider. The Scholar and the Future of the Research Library. A Problem and Its Solution

⁵Shoebbox Machine. Fonte: <http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/speechreco/>

banco de dados relacional⁶, o mesmo em que é baseada a maioria das transações existentes atualmente, como contas bancárias, ações comerciais, reservas de viagens, compras *online* etc.

Em 1975, o Ministério dos Correios e Telecomunicações do Japão realizou um censo⁷ para acompanhar o volume de informações que circulavam no país. Para isso, o número de palavras foi utilizado como unidade de medida em todos os meios de comunicação. O estudo constatou que o fornecimento de informações superava o seu consumo e identificou o crescimento da demanda por comunicação em duas vias.

Em meados da década de 70, surgiram os sistemas MRP (*Material Requirements Planning*), desenvolvidos com o intuito de ajudar empresas a organizar e agendar suas informações. Nessa época, os PCs, que até então eram exclusividade de laboratórios de pesquisas e desenvolvimento e ambientes acadêmicos, estavam começando a ganhar popularidade entre as empresas. Foi nesse período que a *Oracle* lançou a SQL (*Structured Query Language*).

A Primeira Lei de Parkinson diz que os dados se expandem para preencher todo o espaço disponível. Em 1980, a geração de informação era muito maior do que a capacidade de armazená-la. Por isso, uma grande quantidade de dados ficava retida ou era descartada, já que não havia meios de identificar o que era obsoleto e o que era útil (Press, 2013).

Com o avanço da tecnologia, em 1983, as indústrias começaram a se beneficiar de novos métodos para organizar, armazenar e produzir dados, que eram utilizados diretamente para tomada de decisão de negócios.

Com a ascensão dos sistemas MRP, o MRP II (*Manufacturing Resources Planning*) foi desenvolvido com ênfase na otimização de processos e sincronização de materiais com os requisitos de produção. Com o MRP II, era possível gerenciar diversas áreas da empresa, desde o chão de fábrica e a distribuição até os setores financeiro e de recursos humanos. Nessa época, a tecnologia ERP (*Enterprise Resource Planning*) começou a ser adotada por agências governamentais e organizações do setor de serviços (Press, 2013).

A arquitetura para relatórios e análise de negócios, que serviu de base para o armazenamento de dados da IBM, foi desenvolvida por Barry Devlin e Paul Murphy em

⁶Relational Database. Fonte: <http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/reldb/>

⁷Information Society Studies (Routledge Research in Information Technology and Society)

1985, quando surgiu a necessidade de armazenar os dados com alta qualidade, de forma consistente, completa e precisa.

Próximo à década de 90, os sistemas ERP (*Enterprise Resource Planning*) se tornaram mais sofisticados e capazes de coordenar e integrar todas as áreas da empresa, proporcionando acessibilidade, visibilidade e consistência à empresa (Press, 2013).

Em 1989, Howard Dresner expandiu o conceito do termo *Business Intelligence*:

“[...] conceitos e métodos para melhorar a tomada de decisão de negócios, usando sistemas de apoio baseados em fatos.” (Power, 1989)

Em 1992, foi criado o primeiro relatório de banco de dados simples usando o Windows. Estes relatórios permitiram que as empresas criassem um único relatório a partir de uma variedade de fontes de dados, tornando-as habilitadas para empregar inteligência de negócios de forma acessível, o que ajudou a aliviar a pressão da paisagem saturada de dados. A década de 90 foi marcada pelo crescimento do *Business Intelligence* (BI). Os dados gerados por essa tecnologia eram armazenados sob a forma de documentos do *Microsoft Excel* (Press, 2013).

Em 1996, o problema do custo para o armazenamento de dados, causado pelo influxo da informação conseguido com as tecnologias de BI, foi abordado em uma publicação do *IBM Systems Journal*⁸. Como os dados ficaram mais difíceis de serem mantidos, a fim de oferecer mais funcionalidades, o armazenamento digital tornou-se mais rentável do que o de papel.

Ainda em 1996, os fornecedores de ERP enfrentaram um desafio devido à necessidade de redesenhar seus produtos de forma a adotar um negócio colaborativo através da *Internet* de forma contínua (Press, 2013).

Em 1997, o termo Big Data foi utilizado pela primeira vez num artigo da NASA, em que os autores alegaram que o aumento dos dados foi se tornando um problema para os sistemas de computador da época. Isso foi chamado de “Problema de Big Data” (Cox et al, 1997).

Nesse mesmo ano, Michael Lesk⁹ escreveu um artigo sobre o futuro do armazenamento de dados e concluiu que “pode haver alguns milhares de *petabytes* de informações

⁸IBM Systems Journal. Fonte: <http://researchweb.watson.ibm.com/journal/sj/422/morris.html>

⁹How Much Information Is There In the World? Fonte: <http://www.lesk.com/mlesk/ksg97/ksg.html>

ao todo” e que “a produção de fita e disco vai chegar a esse nível até 2000”. O autor afirmou que “seríamos capazes de salvar tudo” e que “nenhuma informação precisaria ser descartada”.

Em 1998, Nic Smith¹⁰ identificou o problema do BI: devido à falta de informação facilmente acessível, homens de negócio não conseguiam obter respostas ou acessar os dados que precisavam para tomar decisões, dependendo cada vez mais de profissionais especializados.

Em 1999, a informação foi quantificada pela primeira vez em um estudo realizado na Universidade de Berkeley¹¹. Nele, os autores conseguiram determinar a quantidade total de informações novas e originais criados no mundo anualmente, bem como sua taxa de crescimento.

Além disso, nesse ano, foi a primeira vez que o termo “*Internet das Coisas*”¹² foi utilizado para descrever o crescente número de dispositivos *online* e o potencial para que eles se comuniquem uns com os outros, muitas vezes sem intervenção humana. E ainda ocorreu a divulgação de um artigo¹³ sobre como escolher e instalar a solução ERP de análise preditiva adequada, para prever como as organizações de todos os tipos fazem negócios.

Em 2001, o termo SaaS (*Software as a Service*) apareceu pela primeira vez em um artigo da *Software Information Industry's (SIIA) eBusiness Division*. Esse termo é um conceito fundamental para muitos dos aplicativos baseados em nuvem que são padrão da indústria atualmente (Press, 2013).

Nesse mesmo ano, Laney (2001) definiu as 3 características que servem de base para o conceito de Big Data: Volume, Velocidade e Variedade.

Em 2002, os principais fornecedores de ERP como SAP, *PeopleSoft*, *Oracle* e *JD Edwards* começaram a utilizar serviços *Web* para conectar suas próprias suítes de forma a tornar mais fácil para o cliente a criação de novas aplicações de dados (Press, 2013).

Em 2005, Tim O'Reilly¹⁴ afirmou que “SQL é o novo HTML” e que “gestão de

¹⁰History of Business Intelligence. Fonte: <http://pt.slideshare.net/nicsmith/history-of-business-intelligence-1236862>

¹¹How Much Information? Fonte: <http://www2.sims.berkeley.edu/research/projects/how-much-info/>

¹²Internet of Things. Fonte: <http://www.rfidjournal.com/articles/view?4986>

¹³Choosing and installing the right ERP solution por Pieter Preston

¹⁴What Is Web 2.0. Fonte: <http://www.oreilly.com/pub/a//web2/archive/what-is-web-20.html>

banco de dados é uma competência central em empresas de *Web 2.0*".

A primeira solução *Open Source* de Big Data foi criada em 2006 para suprir a necessidade de lidar com o grande volume de dados gerado pela *Internet*. O *Hadoop* é uma ferramenta criada para armazenar e analisar grandes conjuntos de dados. Sua flexibilidade permitiu o gerenciamento de dados não estruturados como voz, vídeo, texto etc (C. Anderson, 2008).

Em 2007, foi realizado um estudo para elaborar estimativas e previsões sobre a taxa anual de crescimento de dados digitais. No artigo *The Expanding Digital Universe: A Forecast of Worldwide Information Growth through 2010*, estimou-se que 161 *exabytes* de dados foi criado a nível mundial somente em 2006. Os autores previram que essa taxa iria duplicar a cada 18 meses nos próximos 4 anos. No entanto, a quantidade de dados digitais gerados ultrapassou as previsões, chegando a 1227 *exabytes* em 2010 e 2837 *exabytes* em 2012 (C. Anderson, 2008). No ano seguinte, projetaram que o tráfego de IP dos Estados Unidos poderia chegar a um *zabyte* em 2015 e que a *Internet* desse país seria pelo menos 50 vezes maior do que era em 2006 (Press, 2013).

Em 2008, o conceito de Big Data começou a ser mais discutido nos círculos de tecnologia após a publicação do artigo *The End of Theory: The Data Deluge Makes the Scientific Model Obsolete*¹⁵, em que foram apresentados os impactos positivos e negativos do chamado "dilúvio de dados" (termo utilizado para fazer referência à grande quantidade de dados digitais). Nesse mesmo ano, o termo ganhou credibilidade intelectual após ser mencionado no artigo *Big Data Computing: Creating Revolutionary Breakthroughs in Commerce, Science, and Society*:

"Assim como os motores de busca têm transformado a forma como acessar as informações, outras formas de computação de grande dados podem e irão transformar as atividades de empresas, investigadores científicos, médicos e operações de defesa e inteligência da nossa nação. Computação de Big Data é talvez a maior inovação em computação da última década. Nós apenas começamos a ver o seu potencial para coletar, organizar e processar dados em todas as esferas da vida. Um investimento modesto pelo governo federal poderia acelerar consideravelmente o seu desenvolvimento e implementação." (Manyika et al, 2011)

De acordo com a pesquisa *Gartner EXP Worldwide Survey*, as empresas norte-

¹⁵Big Data Computing: Creating Revolutionary Breakthroughs in Commerce, Science, and Society. Fonte: <http://cra.org/cc/wp-content/uploads/sites/2/2015/05/BigData.pdf>

americanas com mais de mil funcionários armazenam mais de 200 *terabytes* de dados. Por isso, as soluções de *Business Intelligence* tornaram-se prioridade para os CIOs em 2009¹⁶.

O termo *Linked Data*¹⁷ foi usado pela primeira vez em 2009 para fazer referência a um conjunto de melhores práticas para o uso da *Web* de forma a criar ligações entre os dados estruturados. Segundo, Tim Berners-Lee, diretor do *World Wide Web Consortium* (W3C) e inventor da *World Wide Web*, esses dados devem ser publicados explicitamente e devem ter ligação de e para outros conjuntos de dados externos.

Ainda em 2009, Gartner previu que os dados gerados pelas empresas deveriam crescer cerca de 650% nos 5 anos seguintes. Dados esses que representam a mistura de todos os dados operacionais do ERP interno, mais os dados externos que se encaixam às operações da empresa. O *ERP Analytics* surgiu para buscar dar sentido a toda essa informação, incorporando soluções de BI aos ERPs convencionais.

“[...] se um tipo de *Google* da empresa pode apresentar uma forma de reunir todas essas informações não estruturadas em um ambiente baseado em nuvem, e, em seguida, de alguma forma, ligar isso a uma plataforma estruturada, unindo informações estruturadas e não estruturadas, então isso é uma grande, grande coisa.” (Cukier, 2010)

Em 2010, Eric Schimidt, presidente-executivo do *Google*, disse que a quantidade de dados gerada em dois dias equivale à quantidade de dados digitais gerados pela humanidade desde o início da civilização até 2003 (Marr, 2015):

“[...] o mundo contém uma vasta quantidade de informação digital que está ficando cada vez maior, mais rapidamente... O efeito está sendo sentido em todos os lugares, de empresa à ciência, de governos a artes.”

Nesse mesmo ano, empresas como *NetSuite* e *Lawson Software* adotaram o uso de ERP em nuvem¹⁸. Esse tipo de sistema é leve, flexível e acessível, e foi criado para atender a demanda de empresas de médio porte.

Em 2011, o relatório do *McKinley Global Institute* afirmou que, em 2018, haverá um déficit de 140 a 190 mil profissionais capacitados para trabalhar com dados nos EUA.

¹⁶Linked Data - The Story So Far. Fonte: <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>

¹⁷The Future of ERP, Part II. Fonte: <http://www.cio.com/article/2422719/enterprise-software/the-future-of-erp-part-ii.html>

¹⁸Top 10 Trends in Business Intelligence and Analytics for 2011. Fonte: <http://blogs.enterprisemanagement.com/shawnrogers/2011/01/11/top-10-trends-in-business-intelligence-and-analytics-for-2011>

Além disso, o relatório afirmou que questões como privacidade, segurança e propriedade intelectual terão de ser resolvidas antes de o Big Data atingir toda sua capacidade de análise e processamento. Nesse ano, as principais tendências de BI surgiram com a computação em nuvem, tais como: visualização de dados, análise preditiva e Big Data (Marr, 2015).

Segundo estudo *The World's Technological Capacity to Store, Communicate, and Compute Information*, a capacidade do mundo para armazenar informações cresceu a uma taxa anual de 25% entre 1987 e 2007. Em 1986, 99,2% do armazenamento de dados era analógico. Em 2007, 94% dos dados armazenados era digital. Isso representa o quanto a tecnologia evoluiu nos últimos 20 anos. A primeira vez que o armazenamento digital superou o analógico foi em 2002.

Em 2012, o Big Data foi definido no artigo *Critical Questions for Big Data*¹⁹ como:

“[...] um fenômeno cultural, tecnológico e acadêmico que repousa sobre a interação de: (1) Tecnologia: maximizar a potência de cálculo e precisão algorítmica para coletar, analisar, ligar e comparar grandes conjuntos de dados; (2) Análise: com base em grandes conjuntos de dados para identificar padrões, a fim de fazer reivindicações econômicas, sociais, técnicas e legais; e (3) Mitologia: A crença generalizada de que grandes conjuntos de dados oferecem uma forma mais elevada de inteligência e conhecimento que pode gerar *insights* que antes eram impossíveis, com a aura de verdade, objetividade e precisão”.

Em junho de 2012, a *Internet Society* lançou o IPv6 para permitir que produtos e serviços pudessem utilizar esse protocolo permanentemente. Isso para resolver o problema de esgotamento dos 4,3 bilhões de endereços fornecidos pelo formato de endereço IPv4 de 32 bits. *Internet Protocol* versão 6 (IPv6) foi desenvolvido pela *Internet Engineering Task Force* (IETF) e é a versão mais recente do *Internet Protocol* que oferece um sistema de identificação e localização de computadores em redes e rotas de tráfego em toda a *Internet*” (Press, 2013).

De acordo com CIO 2013, os recursos tecnológicos estavam em ascensão:

“(1) Capacidade de executar consultas federadas que dará aos usuários a capacidade de tomar uma consulta e fornecer soluções com base em informações de muitas fontes diferentes, e (2) Capacidade de gerar relatórios de bancos de dados na memória que irá proporcionar um desempenho mais rápido e mais previsível. ” (Olavsrud, 2013)

¹⁹Critical Questions for Big Data. Fonte: <http://www.tandfonline.com/doi/full/10.1080/1369118X.2012.678878>

Em 2013, as empresas começaram a implementar novas tecnologias para analisar e otimizar grandes quantidades de dados para ganhar vantagem competitiva e permanecer relevante no mercado que hoje em dia sofre constantes mudanças (Rivera et al, 2014).

2014 foi considerado o ano da “*Internet das Coisas*” (IoT), uma força poderosa para transformação do negócio. Seu impacto negativo será sentido em todos os setores e áreas da sociedade. Há enormes redes de “coisas” que contêm tecnologia embarcada para sentir ou interagir com seu estado interno ou ambiente externo. De acordo com o infográfico da CSC²⁰, “havia 3,7 bilhões de ‘coisas’ ligadas em uso em 2014 e esse número vai subir para 4,9 bilhões em 2015”. Essa rede de “coisas” permite que cidades inteligentes se tornem realidade. Uma cidade inteligente utiliza análises de contexto de informações em tempo real para melhorar a qualidade e o desempenho dos serviços urbanos, reduzir custos e consumo de recursos, e envolver ativamente os seus cidadãos. Gartner estima que “mais de 1,1 milhões de ‘coisas’ conectadas serão usadas por cidades inteligentes em 2015, incluindo iluminações LED inteligente, monitoramento de saúde, fechaduras inteligentes e várias redes de sensores para coisas como detecção de movimento e controle de poluição do ar”.

A produção de dados está se expandindo a um ritmo surpreendente. Especialistas apontam agora para um crescimento estimado de 4300% na geração de dados anuais até 2020 (B. Marr, 2015).

2.1.3 O que é Big Data

Os fatos descritos no tópico anterior deste trabalho evidenciam que a ideia de analisar dados para entender melhor o que está acontecendo sempre foi explorada pela humanidade, independente do período e da área de aplicabilidade da mesma. O “[...] uso de dados para obtenção de uma informação não é algo novo”, então Big Data não pode ser considerado uma tecnologia nova (Rosa, 2015).

“O Big Data faz parte de uma longa evolução da captação e utilização de dados. Assim como outros desenvolvimentos importantes, como armazenamento digital de dados, processamento de dados e *Internet*, Big Data é apenas mais um passo que vai nos levar à mudança na forma como gerimos os negócios e a sociedade. ” (Marr, 2015)

²⁰Infográfico CSC. Fonte: http://www.csc.com/big_data/flxwd/83638 – *big_data;ust;eginning;to;explode;nteractive;nfographic*

Com a popularização da *Internet*, cada vez mais dados são gerados e esses dados não precisam ser descartados, pois o armazenamento digital é capaz de captá-los e armazená-los, e o Big Data permite analisá-los de uma forma que nunca foi possível antes.

Atualmente, praticamente todas as atividades cotidianas fazem parte do mundo digital, desde compras a ouvir música, tirar fotos e falar ao telefone. Tudo isso deixa um rastro de dados.

“[...] tudo o que fazemos deixa um rastro digital que pode ser usado e analisado. Big Data refere-se aos dados que estão sendo coletados e a nossa capacidade de fazer uso dele. ” (B. Marr, 2015)

Os avanços tecnológicos de *chips* e sensores eletrônicos, da *Internet*, da computação em nuvem e da capacidade de armazenar e analisar os dados mudaram a quantidade de dados que são coletados. Na Era das Grandes Informações, o Big Data analisa informações importantes e relevantes para uma decisão em tempo real e em todas as áreas de negócios.

Outro avanço importante obtido com o Big Data foi a possibilidade de processar e analisar dados não estruturados. Assim, além de números e textos, dados como fotos, *e-mails*, vídeos, gráficos, comportamentos e postagens em redes sociais, gravações de voz e dados de sensor podem ser analisados para encontrar padrões (B. Marr, 2015).

“O Big Data representa volume de informação *versus* velocidade de análise e tratamento do resultado. ” (Barros, 2015)

O primeiro registro documentado do termo Big Data surgiu em 1997, em um periódico escrito por cientistas da NASA, que alegaram que o aumento dos dados estava se tornando um problema para os sistemas de computador da época. A isso deram o nome de “Problema de Big Data” (Cox et al, 1997).

No entanto, a primeira definição para o fenômeno que deu origem ao termo como conhecemos hoje foi em 2001, quando Doug Loney, da empresa *Gartner Group*, apresentou as dimensões do Big Data, os famosos 3 Vs: Volume, Velocidade e Variedade (Laney, 2001).

“Big Data significa coisas diferentes para pessoas diferentes e não há, e provavelmente não terá, um consenso sobre sua definição. Mas o fenômeno é real e está produzindo benefícios em tantas áreas diferentes que faz sentido para todos nós compreendermos seu conceito e sua funcionalidade. ” (B. Marr, 2015)

Por isso, é necessário conhecer algumas definições estabelecidas para o conceito de Big Data, elaboradas por instituições e autores importantes na área. Segue lista elaborada por Rosa (2015) e Canary (2013) com as definições mais relevantes do termo:

- **Oxford Dictionary** (Ano não especificado)

“Conjunto de dados extremamente grandes que podem ser analisados com recursos computacionais para revelar padrões, tendências e associações. ”

- **Kusnetzky** (2010)

“Big Data, em termos simples, refere-se às ferramentas, processos e procedimentos que permitam uma organização criar, manipular e gerenciar grandes conjuntos de dados e seu armazenamento. ”

- **Manyika et al** (2011 apud CANARY, 2013)

“Big Data refere-se a conjuntos de dados cujo tamanho vai além da capacidade das ferramentas de *software* de banco de dados típicos para capturar, armazenar, gerenciar e analisar. ” (Miyazaki, 2015)

- **International Data Corporation** (2011 apud CANARY, 2013)

“As tecnologias de Big Data descrevem uma nova geração de tecnologias e arquiteturas projetadas para extrair economicamente o valor de volumes muito grandes e de uma grande variedade de dados, permitindo alta velocidade de captura, descoberta e/ou análise. ” (Canary, 2013)

- **Dumbill, Howard, et al** (2012)

“ [...] dados que excedem a capacidade de processamento dos sistemas de banco de dados convencionais. Os dados são muito grandes, se movem muito rápido ou não se encaixam às críticas arquiteturas de banco de dados. Para extrair valor destes dados, é preciso escolher um caminho alternativo para processá-los. ” (Dumbill et al, 2012)

- **MCAFEE et al** (2012)

“Big Data como uma forma essencial para melhorar a eficiência e a eficácia das organizações de vendas e *marketing*. Ao colocar Big Data no coração de vendas e *marketing*, os *insights* podem ser aproveitados para melhorar a tomada de decisão e inovar no modelo de vendas da empresa, o que pode envolver utilização de dados para orientar ações em tempo real. ” (Canary, 2013)

- **DEMIRKAN et al** (2012)

“Há o desafio de gerenciar grandes quantidades de dados (Big Data), que está ficando cada vez maior por causa do armazenamento mais barato e evolução dos dados digitais e dispositivos de coleta de informações, como telefones celulares, *laptops* e sensores. ” (Canary, 2013)

- **Mike Phelan** (2012)

“O fenômeno surgiu nos últimos anos devido à enorme quantidade de dados da máquina que está sendo gerado hoje [...] juntamente com as informações adicionais obtidas por análise de todas essas informações, que por si só cria outro conjunto de dados enorme. ” (Canary, 2013)

- **Gartner Group** (2012 apud CANARY, 2013, p.22)

“Big Data, em geral, é definido como ativos de alto-volume, velocidade e variedade de informação que exigem custo-benefício, formas inovadoras de processamento de informações para maior visibilidade e tomada de decisão. ” (Canary, 2013)

- **Taurion** (2013)

“ [...] Big Data = volume + variedade + velocidade + veracidade, tudo agregando + valor. ” (Taurion, 2013)

- **Mayer-Schönberger e Cukier** (2013, p. 4)

“ [...] big data se refere a trabalhos em grande escala que não podem ser feitos em escala menor, para extrair novas ideias e criar novas formas de valor de maneiras que alterem os mercados, as organizações, a relação entre cidadãos e governos etc. ” (Mayer-Schonberger e Cukier, 2013)

- **Davenport** (2014, p.1)

“ [...] big data é um termo genérico para dados que não podem ser contidos nos repositórios usuais; refere-se a dados volumosos demais para caber em um único servidor, não estruturados demais para se adequar a um banco de dados organizado em linhas e colunas; ou fluidos demais para serem armazenados em um *data warehouse* estático. ” (Davenport, 2014)

O conceito que mais se adequa ao intuito deste trabalho é o estabelecido pela IDC, em que se utiliza o termo Big Data para descrever uma nova geração de tecnologias e arquiteturas projetadas para extrair o valor de grandes volumes de dados variados, em alta velocidade, permitindo a captura, a descoberta e/ou a análise desses dados a fim de auxiliar a tomada de decisão.

Apesar de a maioria das definições citadas envolverem a dimensão que corresponde ao volume de dados, é importante que se contemple as três dimensões do Big Data, pois elas são fundamentais para incrementar os resultados das organizações (Rosa, 2015).

Em 2015, Taurion adicionou outras duas variáveis ao Big Data para complementar a relevância do fenômeno e seu efeito ao processo de tomada de decisão. São elas: veracidade e valor. Veracidade porque é preciso não ter dúvidas a respeito do sentido e da autenticidade dos dados, pois uma decisão não pode ser tomada com base em dados duvidosos. E valor porque “[...] é absolutamente necessário que a organização que implemente projetos de Big Data obtenha retorno destes investimentos” (Taurion, 2015).

Segundo Alecrim (2015), o conceito de Big Data “considera não somente grandes quantidades de dados, a velocidade de análise e a disponibilização destes, como também a relação com e entre os volumes”. Para facilitar o entendimento das dimensões e variáveis que fundamentam o conceito de Big Data, este trabalho apresenta a seguir cada uma delas:

a) Volume

É a dimensão que predomina nos conceitos de Big Data, visto que “vem chamando atenção pela acelerada escala em que volumes cada vez maiores de dados são criados pela sociedade” (Taurion, 2015). Apesar disso, o volume “é uma dimensão subjetiva, pois a rápida evolução tecnológica, faz com que o armazenamento e o tratamento de dados

também aumentem” (Rosa, 2015).

Em 2014, a IDC (*Internet Data Corporation*) realizou um estudo para revelar como o surgimento de novas tecnologias contribuem para o aumento do volume de dados. Esse estudo foi baseado em dados gerados por 40 tipos de dispositivos, que permitiram a descoberta de algumas estimativas (Alecrim, 2015):

- a) Em 2013, dois terços dos *bits* do universo digital foram gerados por consumidores, já as empresas foram responsáveis por 85% deste universo;
- b) Em 2013, apenas 22% das informações eram relevantes para análise, sendo que apenas 5% destes foram realmente analisados;
- c) Estima-se que, em 2013, talvez 5% dos dados úteis tiveram valor;
- d) Em 2013, enquanto 40% do universo digital possuía algum tipo de dado que necessitava proteção, apenas 20% destes era protegido de fato;
- e) Em 2013, 20% dos dados estavam em nuvem, sejam eles armazenados (talvez temporariamente) ou processados de alguma forma;
- f) Em 2013, a capacidade de armazenamento disponível conseguia atender apenas 33% do universo digital; e
- g) Em 2014, o universo digital representava 1,7 *megabytes* por minuto para cada pessoa no planeta.

Alecrim (2015) define essa dimensão como “quantidades de dados realmente grandes, que crescem exponencialmente e que, não raramente, são subutilizados justamente por estarem nestas condições”.

“Quando falamos em volume, os números são gigantescos. Se olharmos globalmente, estamos falando em *zetabytes*. Grandes corporações armazenam múltiplos *petabytes* e mesmo pequenas e médias empresas trabalham com dezenas de *terabytes* de dados. Este volume de dados tende a crescer geometricamente e em mundo cada vez mais competitivo e rápido, as empresas precisam tomar decisões baseadas não apenas em palpites, mas em dados concretos. ” (Taurion, 2012)

Para Dumbill, Howard et al, isso representa um desafio imediato às estruturas convencionais de TI, pois reforça a busca pelo armazenamento escalável e a abordagem de sistemas distribuídos (Rosa, 2015).

b) Velocidade

O tratamento dos dados deve ser feito em tempo hábil, geralmente, em tempo real (Alecrim, 2015). Por isso, a velocidade é uma das dimensões que definem Big Data. Avanços tecnológicos como redes de fibra ótica, bandas de telefonia celular, comunicações em tempo real, processamento paralelo e *cluster* de processamento permitiram o aumento da rapidez na troca de dados e informações. Esse aumento seguiu um padrão semelhante ao volume de dados gerados e por isso precisa de tecnologias para seu tratamento (Rosa, 2015).

“[...] na sociedade da informação é crucial saber tratar os dados na velocidade adequada. Dados não tratados e analisados em tempo hábil são dados inúteis, pois não geram informação. Dados passam a ser ativos corporativos importantes e como tal podem e deverão ser quantificados economicamente. ” (Taurion, 2012)

Alecrim (2015) exemplifica a relevância dessa dimensão para Big Data quando diz que “uma operadora de cartão de crédito teria - e causaria - um grande transtorno se demorasse horas para aprovar uma transação de um cliente pelo fato de o seu sistema de segurança não conseguir analisar rapidamente todos os dados que podem indicar uma fraude”.

“[...] é importante saber trabalhar com a velocidade, pois pode ser um limitador da análise, podendo inclusive inviabilizar uma operação, caso um *software* não receba cargas em tempo real, quando a velocidade dos dados é esta. ” (Canary, 2013)

c) Variedade

Essa dimensão se refere aos tipos de dados, que vão desde dados textuais estruturados a dados não estruturados como fotos, vídeos, *e-mails*, *tweets* etc.

“Para soluções Big Data tornam-se necessárias várias tecnologias, desde bancos de dados SQL a *softwares* que utilizem outros modelos, que lidem melhor com documentos, grafos, processamento paralelo etc.” (Taurion, 2012)

Isso ressalta a crescente complexidade do Big Data, pois uma grande quantidade de fontes e tipos de informações são geradas: (a) por dados estruturados entende-se

números, textos, bancos de dados, arquivos sequenciais e com relação de importância; (b) dados semiestruturados, como arquivos XML, são heterogêneos, o que dificulta sua manipulação; (c) e os dados não estruturados são os conteúdos digitais de diversas mídias como vídeos, imagens, áudios, *e-mails*, gráficos etc. Eles representam a maioria dos dados digitais e, por isso, “é necessário saber tratar a variedade como parte de um todo - um tipo de dado pode ser inútil se não for associado a outros” (Alecrim, 2015).

Estruturas relacionais tradicionais não são suficientes para armazenar e tratar dados não estruturados. Por esse motivo foram criados os bancos de dados NoSQL, onde “*No*” significa “*Not Only*”. Isso quer dizer que são utilizadas base de dados SQL e não SQL (Taurion, 2012).

“A importância do *Not Only SQL* explica-se pelo fato do modelo relacional ser baseado no fato que, na época de sua criação, início dos anos 70, acessar, categorizar e normalizar dados era bem mais fácil que hoje. Praticamente não existiam dados não estruturados circulando pelos computadores da época. Também não foi desenhado para escala massiva nem processamento extremamente rápido. Seu objetivo básico era possibilitar a criação de *queries* que acessassem bases de dados corporativas e, portanto, estruturadas. ” (Taurion, 2012)

d) Veracidade

Junto com valor, veracidade é uma das variáveis adicionadas por Taurion ao conceito de Big Data. Segundo o autor, as três dimensões de dados vistas até o momento não são suficientes se não há certeza de que os dados fazem sentido e são autênticos (Rosa, 2015). Os dados precisam ser confiáveis e “é necessário que haja processos que garantam o máximo possível a consistência dos dados” (Alecrim, 2015).

Alecrim (2015) exemplifica a importância dessa variável voltando ao exemplo da operadora de crédito: “Imagine o problema que a empresa teria se o seu sistema bloqueasse uma transação genuína por analisar dados não condizentes com a realidade”.

“A qualidade dos dados e informação é característica essencial para que os usuários interessados (executivos, gestores públicos e a sociedade em geral) usem e (re) usem os dados de maneira apropriada e real, gerando informações críveis para eles mesmos.” (Ribeiro, 2014)

e) Valor

Essa variável não representa apenas valor monetário de retorno. Alecrim (2015) afirma que informação também é patrimônio e que uma solução de Big Data se torna

inviável se o resultado não trouxer benefícios significativos e que compensem o investimento, pois “nenhuma entidade toma decisões e investe se não houver expectativa de retorno” (Alecrim, 2015).

“Todos os dados devem ser valiosos, pois eles passam do uso primário para um potencial uso futuro, afetando seu valor para o negócio e quem tem acesso a eles.” (Mayer-Schonberger e Cukier, 2013)

Essa variável também indica a necessidade de se realizar um estudo aprofundado a fim de diminuir os riscos de um investimento em Big Data antes de efetuar-lo. Para Taurion, o valor da análise de dados só faz sentido se este compensar o custo da sua coleta, armazenamento e processamento (Rosa, 2015).

Segundo Davenport (2014), existem três classes de valor a serem extraídas dos dados: “redução de custos, melhorias no processo decisório e melhorias de produtos e serviços”.

“O valor real dos dados é como um *iceberg* que flutua no oceano: apenas parte é visível no princípio, enquanto boa parte permanece oculta sob a água. As empresas inovadoras capazes de entender isso extraem o valor oculto e aproveitam os enormes benefícios. Em resumo, o valor dos dados precisa ser considerado de todas as maneiras possíveis de uso no futuro, não apenas como são usados no presente. ” (Mayer-Schonberger e Cukier, 2013)

2.2 Aplicações

As técnicas de captação e análise de dados, em volume massivo e em grande velocidade e variedade, vêm se expandindo para novos campos e tende a permear cada vez mais segmentos da sociedade com múltiplas possibilidades de explorar, analisar e tomar decisão a partir do processamento desses dados com algoritmos de Inteligência Artificial. O Big Data está se expandindo para além do mundo da tecnologia e já é realidade em áreas como educação, saúde, *marketing*, serviços etc.

A seguir, são apresentados alguns exemplos reais de aplicações do Big Data, objetivando demonstrar como as técnicas de captação e análise de dados em volume massivo, e em grandes velocidades e variedades, vêm se expandindo para novos campos de atuação da sociedade.

2.2.1 Recursos Humanos

No setor de Recursos Humanos, o Big Data está mudando a forma como as empresas lidam com o capital humano e está substituindo a experiência e a intuição por uma fórmula matemática que aponta com precisão os melhores candidatos a uma vaga de emprego ou os funcionários com maior potencial de crescimento.

Isso é possível porque as ações, interações e conexões das pessoas são registradas digitalmente e geram um grande volume de dados. O emprego de tecnologias que vasculham esses dados para obter informações sobre o desempenho profissional é conhecido como *People Analytics* e está sendo utilizado por empresas de tradição internacional como *Google*, *Coca-Cola* e *General Motors* na gestão de pessoas.

Outro exemplo é a *99 Jobs*²¹, uma empresa de recrutamento que reúne informações sobre empresas e candidatos e promove a apresentação dos mesmos quando uma combinação é encontrada. Das empresas a *99 Jobs* coleta dados como missão, cultura e histórico, e dos candidatos são armazenadas informações relacionadas a objetivos, prioridades e valores. Além disso, a empresa vasculha as redes sociais, buscando pessoas ideais para determinados cargos. Grandes empresas como *Votorantim*, *Microsoft* e *Magazine Luiza* já utilizaram a ferramenta da *99 Jobs* para preencherem vagas em seu quadro de funcionários.

Há também empresas que investiram no Big Data, buscando o aumento de produtividade dos funcionários. É o caso do *Bank of America*, uma das maiores centrais de atendimento telefônico dos Estados Unidos, que adotou o *Media Lab*²². Criada pelo MIT (*Massachusetts Institute of Technology*), essa ferramenta funciona como um crachá inteligente, equipado com um “acelerômetro” para identificar períodos de movimentação, microfones para medir conversas e sensores de proximidade para registrar as interações e os deslocamentos dos funcionários. Cerca de 80 atendentes usaram o crachá por dois meses, permitindo que seu comportamento fosse medido. Os dados armazenados foram processados e analisados. Concluiu-se que empregados que conversavam mais resolviam os chamados mais rapidamente e tinham índices mais baixos de estresse. Baseado nessas informações, o *Bank of America* passou a permitir que mais funcionários tivessem inter-

²¹99 Jobs. Fonte: <https://www.99jobs.com>

²²Media Lab. Fonte: www.media.mit.edu/

valos ao mesmo tempo para estimular as conversas informais. Estima-se que, com isso, o ganho de produtividade chegou a representar em valor real 15 milhões de dólares.

O mérito da análise de dados é fornecer as evidências para apoiar as decisões tomadas por gestores de pessoas. O Big Data não irá substituí-los.

2.2.2 Propaganda e Marketing

A maciça geração de dados na *Internet* possibilita a realização de pesquisas e a geração de relatórios relevantes para o planejamento de *marketing* e vendas sobre o perfil e comportamento de público-alvo de uma empresa. A quantificação de interações, como curtidas, compartilhamentos ou comentários em *blogs*, *sites* de notícias ou redes sociais, permite a identificação de tendências opinativas do momento. Isso pode ser feito por meio de observação ou com processos automatizados, como busca e checagem de palavras-chave na *Internet*. Entender, por exemplo, o que levou o número de visitas a um *site* aumentar em determinado momento e quais assuntos são populares podem representar oportunidades de negócio.

Diante dessa realidade, o Grupo Pão de Açúcar desenvolveu o novo sistema de relacionamento do grupo: o Clube Extra²³. Com ele, o grupo utiliza Big Data para criar ofertas e gerar fidelização à marca, relacionando os consumidores com seus produtos favoritos. Para isso, as informações dos clientes são integradas com as dos fornecedores após estes definirem que tipos de promoções podem oferecer aos consumidores interessados. Assim, a rede começa a oferecer cupons de descontos e exibir anúncios personalizados em seus *sites*.

O Grupo Abril²⁴ também enxergou o potencial do Big Data e lançou a *You-Find Solutions*, que fornece serviços para a análise de dados desestruturados, com o intuito de auxiliar os clientes a compreender o comportamento dos consumidores, prevendo tendências. Para isso, a *startup* utiliza ferramentas de tecnologia que auxiliam na tomada de decisões com base em dados. A ferramenta permite, por exemplo, definir onde implantar um novo ponto de venda, preços a praticar em determinada região, captação e

²³Clube Extra. <http://www.clientesa.com.br/gestao/56788/varejo-brasileiro-em-destaque-no-nrf-2015/ler.aspx>

²⁴Grupo Abril. Fonte: <http://grupoabril.com.br/pt/imprensa/releases/youfind-solutions-nasce-para-atuar-como-consultoria-de-big-data-analytics-com-solucoes-full>

manutenção de clientes, resoluções de gargalos logísticos e testar a aceitação de um novo produto entre os consumidores.

O Big Data permite entender o perfil do consumidor e o seu comportamento nas redes sociais, bem como o hábito de consumo, intenção de compras, relevância e influência.

2.2.3 Localização

A *MapLink*²⁵ é uma empresa de digitalização de mapas e utiliza *softwares* de rastreamento por satélite para calcular o volume de trânsito da cidade de São Paulo. A precisão dos resultados é tanta que a *MapLink* ultrapassou a Companhia de Engenharia de Tráfego (CET) da cidade de São Paulo em termos de confiabilidade.

A CET utiliza câmeras espalhadas pelas principais vias da cidade e a opinião subjetiva de seus fiscais de trânsito para calcular o índice de congestionamento nos horários de pico. Enquanto isso, a *MapLink* monitora os veículos espalhados pela cidade através de sensores de localização de veículos e GPS (*Global Positioning System*), e gera informações em tempo real todos os dias, tais como: condições das vias, períodos estratégicos e tráfego em determinada região.

O método de análise do *software* utilizado pela *MapLink* é baseado na velocidade dos veículos rastreados via satélite pela empresa: se os veículos estão parados, significa que há congestionamento; se estão andando em velocidade baixa, há trânsito; e, por fim, se a velocidade for plena, o fluxo está livre. Com isso, a empresa de localização e roteirização é capaz de determinar aos clientes as condições do trânsito, rotas mais utilizadas e o uso de vias em locais e períodos específicos. Essas informações do trânsito estão sendo utilizadas até mesmo por companhias de logística com o intuito de decidirem os melhores horários e as melhores rotas para suas entregas. Além disso, o armazenamento desses dados permite que órgãos do governo realizem estudos sobre o trânsito para encontrar soluções de mobilidade.

²⁵MapLink. Fonte: <http://maplinkbusiness.com/>

2.2.4 Esporte

A indústria esportiva²⁶ também tem casos de sucesso da utilização do Big Data, pois as ferramentas automatizadas de processamento de dados são capazes de lidar com um grande volume de variáveis e coletar dados de desempenho em tempo real como: frequência cardíaca durante a realização de uma atividade, altura do impulso, queda de força muscular, nível de estafa etc. Isso permite determinar com maior velocidade e precisão quais são os limites físicos de cada atleta, os pontos que devem ser melhorados e até mesmo diagnósticos táticos como erros de fundamento e falha de posicionamento. Além disso, o monitoramento de dados possibilita o estudo das características de jogo ou de comportamento do adversário, o que garante maior eficácia na preparação para o jogo ou competição. Por isso, modalidades como vôlei, basquete, tênis, artes marciais e futebol adotaram técnicas de análise e processamento de dados, buscando melhores resultados, melhor rendimento e até mesmo definir estratégias de jogo.

Uma das pioneiras na utilização do Big Data no esporte, a seleção brasileira de vôlei instalou diversas câmeras em pontos estratégicos das quadras no Centro de Desenvolvimento de Voleibol, em Saquarema no Rio de Janeiro, para coletar imagens dos treinamentos dos atletas durante a preparação para as competições. Além disso, equipamentos eletrônicos são presos ao corpo dos atletas, permitindo o monitoramento completo do corpo do atleta durante as atividades. São coletados dados como a frequência de batimentos cardíacos, a avaliação de contrações e distensões musculares em cada movimento, os níveis de impulso, o tempo de resposta dos jogadores diante das jogadas e a movimentação coletiva em quadra. Todos esses dados são processados e geram informações sobre o rendimento e o desempenho tático dos atletas, o que facilita a criação de treinamentos direcionados aos pontos fracos identificados após a análise dos dados.

A conquista da Copa do Mundo de 2014 pela Alemanha se deve muito ao Big Data. Antes mesmo de a competição começar, a comissão técnica do time alemão utilizou um sistema de coleta, agregação e processamento de dados para ter acesso a informações das equipes adversárias como jogadas ensaiadas, movimentações, pontos fortes e fraquezas. E, com isso, conseguiu tornar os treinamentos e a preparação dos jogadores mais eficazes.

²⁶Big Data no Esporte. Fonte: <http://bigdatabusiness.com.br/como-solucoes-de-big-data-podem-ajudar-atletas-e-esportistas/>

Soluções de Big Data têm sido aplicadas também para otimização de performance dos atletas. É o caso do Bayern de Munique, time de futebol da Alemanha, que busca com as técnicas compreender com perfeição características, como a energia cinética envolvida nos movimentos de seus jogadores durante os treinamentos. O sistema pode, por exemplo, avisar ao preparador físico que um jogador já sofreu uma lesão na última vez que fez determinado exercício. Com isso, a rotina de treinamentos pode ser adaptada de acordo com as limitações ou necessidades de cada atleta, reduzindo lesões. No Brasil, o Grêmio é um dos clubes que se beneficia das tecnologias de análises de grandes dados com o intuito de realizar o diagnóstico completo das limitações de cada atleta, buscando otimizar o desempenho dos mesmos.

No MMA (*Mixed Martial Arts*), as soluções automatizadas fornecem análises descritivas, preditivas e de diagnóstico aos atletas e seus treinadores. São geradas informações acerca do treinamento dos atletas e das performances dos adversários. Com isso, o treinamento é programado para que o atleta esteja preparado para praticamente todas as ações do adversário, reduzindo a imprevisibilidade do octógono.

2.2.5 Segurança

O setor de segurança também está usando o Big Data a seu favor para combater e prevenir crimes. Há diversas ferramentas, baseadas em modelos matemáticos e no alto poder de processamento, que oferecem meios para descobrir padrões, locais mais perigosos e comportamentos similares de criminosos, entre as quais se destacam: Detecta, utilizada pela a polícia de São Paulo; *Blue Crush* e *Domain Awareness System* (DAS), soluções da *Microsoft* que auxiliam no combate e na prevenção de crimes nos EUA.

- Detecta

Em agosto de 2014, a Secretaria de Segurança Pública de São Paulo, em parceria com a *Microsoft*, lançou o sistema Detecta. Trata-se de uma ferramenta de Big Data que tem por finalidade ajudar policiais e autoridades a verificar o histórico de ocorrências de um lugar específico. Para isso, o sistema se conecta às câmeras do governo e a um banco de dados com informações extraídas das centrais telefônicas 190 (Polícia Militar) e

193 (Corpo de Bombeiros). A proposta do Governo de São Paulo é aumentar o número de câmeras conectadas ao sistema com a inclusão das câmeras do setor privado. Isso já acontece desde 2012 em Nova York, onde se utiliza um sistema similar ao Detecta.

Quanto mais bancos de dados forem conectados ao sistema, mais precisas serão as informações. Dessa forma, policiais equipados com dispositivos móveis, como *tablets* ou *smartphones*, conseguem acessar informações em tempo real e ter, numa determinada operação, o que se chama de Consciência Situacional. Assim, ao atender uma ocorrência, por exemplo, o policial recebe mais informações sobre a situação que vai encontrar, como: que outros crimes ocorreram na área recentemente; que suspeitos de outros crimes residem nas proximidades; se foi registrada apreensão de armas no local etc. Essas informações podem envolver desde características de um suspeito, coletadas a partir de câmeras de segurança, até a forma como determinado crime foi cometido. Também é possível conferir imagens do local do crime que já foram registradas no sistema antes de o policial chegar no mesmo.

O Detecta é capaz de receber informações de qualquer tipo de sensor eletrônico, como câmeras e leitores de placas de carros. Essas informações geram *feeds* que são inseridos em um banco de dados. A partir do relacionamento desses *feeds*, utilizando Big Data com *Business Intelligence* (BI), os alertas são criados em tempo real e enviados automaticamente para o usuário. O sistema funciona como um buscador, indexando dados policiais e fazendo associações entre eles, e é multiplataforma. Isso significa que o usuário pode estar usando um computador, um *smartphone* ou um *tablet*, e poderá acessar o sistema, com comando a partir do Copom (Centro de Operações da Polícia Militar), do Cepol (Centro de Comunicações e Operações da Polícia Civil) e Departamento de Inteligência (Dipol).

Em São Paulo, a polícia e os bombeiros recebem cerca de cento e oitenta mil ligações por dia. Cada ligação dessa pode se dividir em centenas de ações, tecnicamente chamadas de transações. Além disso, todo *tablet* e aparelho de GPS pode enviar uma localização a cada dez ou quinze segundos. São milhões de transações geradas diariamente. Sem Big Data é praticamente impossível relacionar todas essas informações corretamente para que elas se tornem alertas para o usuário do Detecta.

- *Blue Crush*²⁷

Em uma iniciativa da prefeitura de Memphis, em parceria com o Departamento de Criminologia da Universidade de Memphis e com a IBM, o *Blue Crush* (*Crime Reduction Utilizing Statistical History*) foi criado para coletar dados das ocorrências policiais, dos padrões dos criminosos e dos locais dos crimes, permitindo maior precisão da polícia no combate ao crime.

Para isso, todos os oficiais da cidade ganharam PDAs (*Personal Digital Assistant*) para registrar os fatos assim que ocorriam. Os dados gerados eram analisados pelo *software* SSPS, programa de análise preditiva da IBM, e processados pelo programa GIS, um serviço de mapeamento da companhia ESRI, para entregar um mapa com os locais prováveis de novos crimes. Com os dados em mãos, a polícia passou a patrulhar essas áreas e conseguiu reduzir em cerca de 25% os crimes na cidade no período de 2006 a 2010.

A iniciativa continua e ainda ganhou o reforço de um centro integrado de segurança chamado *Cyberwatch*, que distribui mapas e informações sobre crimes eletronicamente aos cidadãos, incluindo informações e imagens de indivíduos procurados.

- *Domain Awareness System* (DAS)²⁸

Desenvolvido pela *Microsoft*, o *Domain Awareness System* (DAS) é um sistema que processa dados telefônicos do 911 - número que corresponde ao serviço de emergência nos Estados Unidos - e imagens de câmeras de segurança públicas e privadas. O sistema é capaz de reconhecer rostos, placas de carro e radiação nuclear, e permite ainda localizar o policial mais próximo a cada ocorrência.

Esse sistema é utilizado na cidade de Nova York desde 2012, quando a prefeitura lançou um mapa interativo do crime na cidade, que permite filtrar por bairro e rua, além de verificar o tipo de crime ocorrido. Esse mapa auxilia os cidadãos e os investigadores a identificarem onde os delitos e os crimes violentos ocorrem com maior frequência. O DAS foi usado como modelo para a implantação do *Detecta* em São Paulo.

²⁷Blue Crush. Fonte: <https://fcw.com/articles/2012/11/02/big-data-memphis.aspx>

²⁸DAS. Fonte: <http://exame.abril.com.br/tecnologia/noticias/big-data-contra-o-crime>

2.2.6 Economia

A economia utiliza números para provar ou refutar teses, gerando bancos de dados cada vez maiores e mais complexos, capazes de agregar informações valiosas sobre o comportamento das pessoas. Isso permite a análise do comportamento econômico com mais critério.

Geralmente, os dados econômicos são defasados e isso dificulta a formulação de políticas econômicas. O padrão de pesquisas na *Internet* pode antecipar e refletir condições econômicas, prevendo, inclusive, movimentos futuros dos preços em alguns mercados financeiros. Isso porque há uma conexão entre o que as pessoas pesquisam na *Internet* e seu comportamento posterior. Com o Big Data, é possível aplicar algoritmos de análises preditivas para obter informações úteis e confiáveis a partir das buscas realizadas na *Internet*. Por exemplo, mais pessoas pesquisando informações sobre automóveis pode preceder um aumento nas vendas de automóveis.

Recentemente, um estudo da *Cornell University Library*, chamado *Social media fingerprints of unemployment*²⁹, descobriu padrões universais da atividade humana e da comunicação interpessoal. Nesse estudo, foram investigados os desvios desses padrões para revelar informações sobre a situação socioeconômica de algumas regiões. O estudo encontrou uma relação entre o conteúdo de *tweets* e o desemprego. Para isso, foram quantificadas características individuais de comportamento de mais de 145 milhões de mensagens distribuídas em mais de 340 regiões econômicas diferentes e concluiu-se que regiões com maior diversidade de fluxos de mobilidade, hábitos diurnos e estilo gramatical correto exibem taxas de desempregos menores. Os resultados mostram que os indicadores econômicos de custo-benefício podem ser construídos com base em conjuntos de dados publicamente disponíveis de mídia social.

O monitoramento das condições econômicas também se tornou possível com a coleta de metadados telefônicos. Um estudo realizado por economistas de Havard e do MIT (*Massachusetts Institute of Technology*)³⁰ coletou padrões telefônicos de 10 milhões de usuários de telefone celular em um país europeu não identificado que estava sofrendo com altas taxas de desemprego. Os dados foram colocados para interagir em um mo-

²⁹Social media fingerprints of unemployment. Fonte: <http://arxiv.org/abs/1411.3140>

³⁰Tracking employment shocks using mobile phone data. Fonte: <http://rsif.royalsocietypublishing.org/content/12/107/20150185>

delo matemático que levava em consideração fatores como rede de contatos e ligações realizadas e recebidas, e se mostraram tão precisos e eficientes para monitorar o desemprego quanto pesquisas tradicionais realizadas por amostragem. Isso significa que o Big Data está permitindo que novas fontes de dados sejam utilizadas para definir e monitorar comportamentos econômicos.

2.2.7 Dados climáticos

Ferramentas de Big Data estão sendo utilizadas para definir previsões do tempo com mais precisão. Para isso, é coletada uma grande quantidade de dados meteorológicos, geográficos e históricos do local. Esses dados são gerenciados, separados e analisados pela *Deep Thunder*³¹, tecnologia desenvolvida pela IBM, que retorna vídeos e animações bidimensionais com padrões de chuva e movimentação de nuvem.

Ao contrário do que é comumente usado em sistemas meteorológicos tradicionais, esse sistema se restringe ao padrão de previsão de tempo da área analisada, o que oferece uma precisão local dos resultados, capaz de prever as condições climáticas com até três dias de antecedência. Apesar de não ser perfeita, essa tecnologia diminuiu em até 50% a taxa de erros nas previsões meteorológicas.

O Big Data também tem sido aplicado em soluções de energias renováveis, como a solar e a eólica. Devido à intermitência dessas fontes de energia, é preciso haver uma grande capacidade de reserva para garantir a continuidade do fornecimento de energia. Por isso, pesquisadores do Centro Nacional de Pesquisa Atmosférica (NCAR) passaram a coletar, 24 horas por dia, dados reais de ventos que passam pelas unidades de aerogeradores da empresa. Essas informações são cruzadas com centenas de fontes de informações climáticas disponíveis para gerar prognósticos de ventos. Isso reduz a necessidade de utilizar fontes reservas de energia como a termelétrica e a hidrelétrica.

Com o Big Data, é possível analisar um volume grande de dados, permitindo que os modelos sejam aprimorados. Assim, quanto mais cresce a instalação das fontes de energias renováveis, maior é a precisão da previsão dos sistemas.

³¹Deep Thunder. Fonte: <http://www.computerworld.com/article/2488618/big-data/big-data-key-to-bringing-hyperlocal-weather-forecasts-to-georgia-farmers.html>

2.2.8 Comercial

A utilização do Big Data já é realidade também nos setores comerciais de grandes empresas que buscam potencializar suas vendas. A *Walmart* é considerada a maior varejista do mundo e se tornou referência por conseguir coletar dados *online* para impulsionar as vendas de suas lojas físicas. Os *softwares* desenvolvidos pela empresa monitoram quando a discussão sobre o campeonato de futebol americano se intensifica na *Internet* em diferentes cidades dos Estados Unidos. Diante disso, os gerentes das lojas dessas regiões organizam as vitrines de forma a expor produtos de determinados times.

Em consequência da demanda interna por soluções de Big Data, a varejista criou a *WalmartLabs* em 2011, uma subsidiária de tecnologia que nasceu da compra da *Kosmix*, uma *startup* americana que ganhou notoriedade no mercado em 2010, ao desenvolver um sistema que ajuda no gerenciamento de estoques das varejistas durante a *Black Friday* (sexta-feira negra), principal dia de queima de estoques do varejo nos Estados Unidos. O *software* da *Kosmix* detecta, pela localização dos celulares dos clientes, o número de pessoas em cada loja. Com essa informação, os estoques de unidades que estão com vendas em baixa são enviados para as que estão vendendo mais.

No Brasil, a Renner investiu em um sistema que compara as vendas de suas lojas em tempo real. Isso tornou possível a identificação de números fora do padrão. Por exemplo, nos dias em que as unidades localizadas em lugares frios vendiam muitos casacos, os gerentes das lojas com desempenho abaixo da média recebiam um aviso do sistema para mudar a posição do produto na vitrine.

Recentemente, o projeto foi aprimorado para prever o estoque necessário de cada produto mediante a aceitação do mesmo no *Facebook*, e gerar relatórios que consolidam dados como a previsão do tempo e comentários em redes sociais. Além de reunir essas informações, o sistema seleciona o que é confiável.

2.2.9 Saúde

Um exemplo da utilização de Big Data no setor de saúde é o *HealthMap*³², uma iniciativa do Hospital Infantil de Boston que reuniu uma equipe multidisciplinar com pesquisadores,

³²HealthMap. Fonte: <http://www.healthmap.org/pt/>

epidemiologistas e desenvolvedores de *software* para mapear em tempo real a ocorrência de determinadas doenças em países emergentes.

A ferramenta submete diversas fontes de dados informais da *Internet* a algoritmos de Inteligência Artificial para filtrar, dentre o grande volume de dados armazenado, um pacote de informações fácil de usar. Os dados obtidos são disponibilizados gratuitamente no *site* da ferramenta e servem de base para bibliotecas, departamentos de saúde, governos e viajantes internacionais.

HealthMap reúne diferentes fontes de dados, desde agregadores de notícias *online* e redes sociais a relatos de testemunhas, fóruns especializados em infectologia e relatórios oficiais de órgãos públicos, para alcançar uma visão unificada e abrangente do estado atual de doenças infecciosas e os seus efeitos sobre a saúde humana e animal.

O *software* contém filtros que identificam e descartam dados irrelevantes, além de classificar as informações que interessam e dispor as doenças em um mapa. As epidemias são categorizadas em diferentes cores, de acordo com sua intensidade, e é possível também localizar as que estão mais próximas, ou então acessar diversas notícias sobre o assunto, além de várias outras opções de personalização dos dados. Surtos como os de dengue e febre amarela podem ser visualizados em nove línguas diferentes.

A ferramenta facilita a detecção precoce de ameaças globais para a saúde pública. Um exemplo da eficiência no tratamento dessas informações foi o alerta emitido pelo *site* sobre a disseminação do ebola no continente africano em 2014. O alerta foi emitido nove dias antes de a Organização Mundial de Saúde (OMS) ser notificada oficialmente pelo Ministério da Saúde de Guiné sobre a natureza da doença. Na ocasião, a ferramenta identificou um surto de febre hemorrágica na região e encontrou em sua base de dados menções ao ebola, lançando um alerta.

A ferramenta está sendo utilizada para um estudo sobre o vírus Zika que busca se antecipar à propagação internacional do vírus encontrado no Brasil. O *HealthMap* conta com uma versão adaptada para dispositivos móveis chamada *Outbreaks Near Me* (Epidemias Próximas a Mim), que está disponível para os sistemas Android e iOS.

2.2.10 Turismo

Reservas de companhias aéreas, hospedagem em hotéis, aluguel de carros e reserva ou compra *online* de passagens deixam um rastro de dados na *Internet* ³³. Isso também ocorre com as curtidas em redes sociais e com as críticas enviadas às empresas pelo Canal de Atendimento ao Cliente, ou quando os preparativos para viagens são discutidos *online* em *blogs* e *sites* de ranqueamento especializados no assunto. Com isso, é gerado um grande volume de dados que pode ser utilizado para obtenção de uma visão mais integrada do segmento de Turismo, o que oferece às companhias de viagens uma oportunidade para melhorar os seus processos atuais, impulsionar a inovação e construir relações melhores com seus clientes.

O Big Data permite que os dados coletados, sejam eles estruturados ou não-estruturados, sejam processados e analisados para ajudar os agentes de viagens a tomar decisões e agir sobre os dados em tempo real. Dessa forma, as companhias de viagens conseguem identificar novas oportunidades de ofertas de produtos e serviços, cada vez mais assertivos para públicos cada vez mais específicos, e com isso, oferecer uma experiência mais inteligente de viagem.

³³Big Data no Turismo. Fonte: <http://convergecom.com.br/tiinside/services/20/08/2013/big-data-uma-revolucao-necessaria-para-a-industria/>

3 Big Data nas Escolas

Há um interesse crescente por parte das instituições de ensino para tirar proveito do Big Data a fim de melhorar o desempenho dos alunos, aumentar a eficácia do professor e reduzir a carga de trabalho administrativo (Schmarzo, 2014).

A aplicação de soluções de Big Data para o processamento e a análise de dados educacionais para personalizar o aprendizado denomina-se Ensino Adaptativo e sua relevância está na “capacidade de processar massas de dados em escala, através da análise e da comparação de comportamento de milhares de estudantes para gerar conhecimentos generalizáveis sobre o processo de aprender e as dificuldades inerentes a determinados conteúdos ou estratégias pedagógicas” (Scaico et al, 2014).

Dados sempre fizeram parte da rotina escolar e são necessários para a tomada de decisões, tanto de gestores quanto de professores e coordenadores pedagógicos. Através do processamento e da análise desses dados, a tecnologia possibilita a obtenção de informações detalhadas sobre interesses, habilidades, dificuldades, nível de aprendizado e até mesmo alguns hábitos e características emocionais de alunos ou escolas inteiras. Além disso, permite que os professores entendam melhor as turmas e obtenham um panorama mais exato do nível de conhecimento dos estudantes (Sasaki, 2015).

O Big Data oferece variedade, quantidade, exatidão e rapidez. Isso significa que é possível analisar e processar grandes volumes de informações de tipos variados e obter resultados exatos em tempo hábil para tomada de decisão.

“Uma das formas de obter esses dados é por meio de avaliações externas, que, associadas a tecnologias educacionais, geram relatórios de aprendizado bem detalhados tanto dos alunos individualmente quanto de toda uma escola. Quando analisados e contextualizados por professores e gestores, essa informação pode nortear desde o desenvolvimento de planos de aula que atendam melhor às necessidades mais imediatas de uma classe até toda a política pedagógica da escola.” (Sasaki, 2015)

No tópico seguinte deste trabalho, são descritos alguns conceitos do termo Ensino Adaptativo e como o Big Data está sendo utilizado na educação para aprimorar os modelos educacionais. Em seguida, são apresentadas algumas ferramentas de Ensino Adaptativo

para exemplificar a aplicação de Big Data na Educação.

3.1 Conceito e prática do Ensino Adaptativo

O uso de plataformas *online* de apoio ao processo de aprendizagem gera uma grande quantidade de dados automaticamente. A partir da análise desses dados, é possível levantar uma série de informações sobre os estudantes, como as dificuldades que apresentam, o formato que aprendem melhor, o nível de interações, horários mais produtivos etc. Com isso, é possível personalizar o ensino e focar nas diferenças de cada um, melhorando o processo de aprendizado, tornando-o mais eficaz e atraente.

A utilização do Big Data na Educação denomina-se Ensino Adaptativo, representa uma tendência na área educacional mundial e consiste em métodos de aprendizagem que utilizam computadores como dispositivos de ensino interativo que se adaptam de acordo com as necessidades específicas de cada aluno, mediante suas respostas a perguntas, tarefas e experiências. Nesse modelo, cada professor consegue ser um tutor particular para toda uma sala de aula, o que torna o sistema de ensino personalizado mais inclusivo.

“[...] o uso de Big Data no espaço de aprendizagem *online* dará às instituições as ferramentas de previsão que precisam para aprimorar os resultados de aprendizagem para os alunos individualmente.” (Ferreira, 2014)

O Big Data permite coletar e analisar em “tempo de aula” uma infinidade de dados, históricos e atuais, sobre o desempenho dos alunos, áreas críticas e outras informações que permitam adaptar estratégias e métodos de ensino para melhorar os resultados dos alunos. Isso significa que a utilização das técnicas de Big Data possibilita

“capturar muitas informações sobre os estudantes e sobre a sua interação com os conteúdos instrucionais, ambientes de aprendizagem e sobre o processo de avaliação, que são difíceis de ser coletadas e interpretadas por meio de abordagens tradicionais que não façam uso de tecnologias com alto poder de processamento e análise”. (Scaico et al, 2014)

O Ensino Adaptativo busca aperfeiçoar o processo de aprendizagem, auxiliando os alunos a construírem seus próprios percursos de aprendizagem, através de recursos tecnológicos para educação *online*. Vários sistemas educativos³⁴ desse tipo foram implementados, tais como: hipermídia adaptativa educacional, sistemas de tutores inteligentes,

³⁴Aprendizagem Adaptativa. Fonte: https://pt.wikipedia.org/wiki/Aprendizagem_adaptativa

testes de adaptação computadorizada, agentes pedagógicos baseados em computador, entre outros.

Com essas ferramentas, busca-se implementar o

“processamento dos dados de cada aluno quanto ao seu rendimento, notas, leituras, vídeos assistidos, horas dedicadas ao estudo e todos os referenciais que de algum modo indiquem quem é este estudante, porque vai muito bem ou muito mal nos estudos, quais são suas perspectivas futuras quanto à educação, carreira e vida” (Machado, 2015).

E, com base nesses dados, é possível ajudar alunos que têm rendimento baixo a melhorar suas notas, a aprender a estudar, a ter mais foco, a dedicar-se mais, a ter metas plausíveis e, até mesmo, a identificar eventuais dificuldades de aprendizagem neles presentes (Machado, 2015).

“[...] o Big Data pode ser usado por plataformas educacionais adaptativas, que coletam informações ligadas às rotinas de estudo do aluno, o nível de proficiência em cada matéria, seus interesses e habilidades. A partir delas, a ferramenta elabora um plano personalizado de estudos, recomendando exercícios e aulas que trabalhem deficiências específicas e evitando conteúdos que o aluno já domine completamente, o que ajuda a fazer com que se sinta sempre desafiado.” (Sasaki, 2015)

Além disso, busca-se oferecer referenciais para estudo que sejam adequados aos padrões e ritmo do aluno, o que significa oferecer os conteúdos em mídias diferentes, que se adequem melhor a sua capacidade e condição de leitura de mundo. Isso porque “há estudantes que aprendem mais vendo vídeos sobre um determinado tema, outros são mais afeitos a leitura de textos, há aqueles que entendem melhor vendo uma simulação ou animação” (Sasaki, 2015).

Para estabelecer um nível de inferência para atestar o que um estudante é capaz de fazer e o que ele aprendeu, é preciso acumular e observar dados individuais e compará-los a dados de outros alunos. Isso para que seja possível classificar e agrupar estudantes que possuem o mesmo perfil e habilidades ou que enfrentam as mesmas dificuldades e, com isso, estabelecer uma maneira eficiente de articular atividades e intervenções pedagógicas para estimular ações para a melhoria da aprendizagem, do engajamento e da retenção dos alunos, além de melhorar o que se conhece sobre a relação ensinar e aprender (Scaico et al, 2014).

Scaico et al (2014) lista uma série de questões complexas que podem ser respondidas com o uso do Big Data na Educação: “o que um estudante sabe? O que ele não

sabe? Até que ponto um conteúdo ou uma estratégia pedagógica pode funcionar para um grupo de estudantes? Que sequência de assuntos pode ser mais adequada para um estudante com um perfil específico? Que resultados de aprendizagem podem ser medidos quando um estudante interage com um determinado tipo de conteúdo? Até que ponto uma avaliação realmente atingiu a medição a que se propunha? Como ter certeza sobre o nível de proficiência de um aluno sobre um determinado conteúdo curricular? A falha do aluno ocorreu porque ele não aprendeu ou ela se deveu à distração, esquecimento ou a uma pergunta mal formulada? Como aumentar as chances de um estudante obter sucesso em uma avaliação?”

Os benefícios trazidos pelo uso do Big Data na Educação não são exclusivos dos discentes. Ao dispor de mais conhecimentos sobre o processo de ensino e aprendizagem, as instituições podem melhorar seus modelos de gestão de aprendizagem e de avaliação, construir novas práticas de ensino, aumentar a discussão sobre as políticas educacionais existentes e promover novas experiências de aprendizagem, capazes de melhorar a maneira como os professores ensinam, os estudantes aprendem e a escola funciona. Além disso, pode “conduzir mais eficientemente o planejamento de cursos, apontando como um conteúdo pode ser melhor dividido em tópicos, como estes devem ser sequenciados e que configurações podem gerar curvas maiores de aprendizagem” (Scaico et al, 2014).

De acordo com Scaico et al (2014), o Big Data pode ser utilizado pelas instituições de ensino para coletar informações como:

- a) Dados relacionados à identidade dos usuários: quem são eles, quais são as suas permissões e de onde o acesso está sendo realizado;
- b) Dados relacionados à interação do usuário com o ambiente de aprendizagem: se referem à experiência que os estudantes tiveram ao utilizar o ambiente. Dados dessa categoria incluem métricas de retenção, taxa de cliques, páginas visitadas e taxas de retorno;
- c) Dados inferidos sobre os conteúdos: se referem à compreensão de evidências que sejam capazes de apontar o nível de eficiência existente na relação entre um conteúdo escolar e os ganhos de aprendizagem de estudantes ou grupos de estudantes. Al-

cançar métricas para gerar este tipo de dados é mais difícil por exigir a exposição de grupos de estudantes a diferentes conteúdos e posterior avaliação, que pode resultar em evidências que apontem a eficiência dos conteúdos;

- d) Dados relacionados aos sistemas de informação disponíveis nas instituições de ensino: se referem aos dados de natureza mais burocrática, que já são, em grande parte, coletados pelos mecanismos tradicionais da escola. Envolve notas, registros disciplinares e de assiduidade. Em pequena escala, este tipo de dados é de pouca utilidade, já que há pouca informação dessa natureza por cada estudante. Em larga escala, os dados podem assistir à geração de recomendações para a escola como um todo, apesar de ainda se constituírem dados superficiais que podem não ser confiáveis ou de pouca utilidade para a tomada de ações, especialmente pelo fato de apontarem problemas sociais difíceis de serem resolvidos no contexto escolar;
- e) Dados de inferência sobre os estudantes: são os dados mais difíceis de serem gerados porque dizem respeito ao comportamento de cada aluno diante do processo de aprendizagem de um conteúdo, que pode ser afetado em diferentes momentos por diferentes fatores e pelos recursos instrucionais utilizados. Dados dessa natureza procuram inferir o que um aluno aprendeu, qual é o seu comportamento e como isso influencia a sua aprendizagem, quais são as suas motivações e interesses, dentre outros aspectos.

Ferreira (2014) cita o uso de Big Data para prevenir a evasão a escolar:

“O Big Data poderia prever os fracassos e incorporar novas dinâmicas ao aprendizado que contornassem os obstáculos de cada indivíduo, numa espécie de alfaiataria educacional baseada em dados.”

Faria (2015) sinaliza para uma falha que ocorre constantemente nos modelos tradicionais de educação que não levam em conta que “cada pessoa é única e possui maneiras próprias de compreender o mundo. Seja através da leitura, da música, da escrita ou de jogos, cada aluno cria seus próprios mecanismos para aprender”. Nesses modelos, as avaliações são aplicadas em massa para todos os alunos, da mesma forma, “sem considerar as características próprias de aprendizagem e aplicação do conhecimento de cada um. Com o auxílio do Big Data, torna-se possível medir o desempenho dos alunos com foco nos *gaps* de conhecimento e assim potencializar os resultados”.

“[...] o Big data teria efeito revolucionário por não apenas transmitir informação, mas por ser capaz de ensinar com base nos nossos comportamentos e cliques assim como os anúncios do *Facebook* e do *Google*. ” (Ferreira, 2014)

Segundo Bettocchi (2015), o Ensino Adaptativo está baseado nas seguintes premissas:

- As pessoas possuem diferentes ritmos de aprendizagem, os quais podem inclusive variar de acordo com o assunto. Por exemplo, maior facilidade em aprender Matemática do que Português
- Diferentes pessoas reagem melhor a diferentes formas de apresentação de conteúdo (animação, palestra, diagramas, texto escrito) e tarefas (exercícios de múltipla escolha, pesquisas, análises)
- Um professor tem grande dificuldade de atender a essas demandas individuais dos alunos em uma aula presencial com turmas de trinta alunos ou mais

Sendo assim, a proposta do Ensino Adaptativo é complementar as aulas presenciais, em que são trabalhados debates e outras ações coletivas, com percursos individuais de aprendizagem que são elaborados por meio das escolhas do aluno, de um *software* que acompanha o seu desempenho e, quando possível, de tutores (Bettocchi, 2015).

“Ao incorporar o Big Data como ferramenta de análise de dados no ambiente escolar, é possível desenvolver métodos mais eficazes de ensino, buscando-se não apenas facilitar a aprendizagem do aluno como também tornar o trabalho do professor menos maçante e mais dinâmico, como um orientador, um *coach* que mostra o caminho e acompanha o aluno direcionando suas estratégias até a conquista dos seus objetivos. ” (Faria, 2015)

O modelo mais comum é aquele no qual o aluno escolhe um assunto para estudar (Matemática, Português, História), faz um teste para determinar seus pontos fracos e a partir daí vai assistindo aulas em vídeo e realizando as tarefas propostas. Em seguida, o sistema indica quais pontos devem ser revistos, propõe outras tarefas e indica aulas de reforço, para que o aluno possa construir seu conhecimento. Recursos como mensagens de elogio e/ou técnicas de “gameificação” (consiste em utilizar os princípios dos jogos que fazem o usuário viciar em jogar, como estipular objetivos fáceis de serem alcançados, mas

não óbvios ao aluno, dar pontos quando ele acerta, indicar seu progresso em uma barra) que traçam objetivos e indicam o avanço do aluno também são utilizados a fim de evitar a evasão. Atualmente, esse é o principal uso das plataformas de Ensino Adaptativo no Brasil (Bettocchi, 2015).

No entanto, em um futuro próximo, planeja-se conjugá-las com algoritmos de Big Data e Data Mining para que essa personalização das técnicas de ensino para o aluno fique ainda mais refinada (Bettocchi, 2015).

“No futuro, com a convergência do Ensino Adaptativo e do Data Mining (exploração e análise de bases de dados), será possível explorar correlações incríveis entre aprendizagem e conteúdo, permitindo que atuemos sobre ambos de forma quase científica. Para facilitar a adaptabilidade dos assuntos estudados, os conteúdos passarão a ser gerados sempre em pedaços e em diferentes formatos como texto, vídeo e áudio.” (Campos, 2014)

Isso permite que planejamentos curriculares sejam realizados a partir de conclusões como:

- Alunos que tiveram dificuldade com as questões de redação do Enem costumam ter problemas na disciplina X do 3º período do curso Y, portanto é preciso planejar uma revisão de técnicas de redação nessa disciplina se tivermos muitos alunos com esse perfil no próximo ano (Bettocchi, 2015).
- Alunos que tem dificuldade na matéria X do terceiro ano do curso, geralmente foram fracos na prova Y do vestibular. Isso permitiria que a faculdade e seus professores se preparassem de antemão para cuidar de alunos com baixo rendimento naquela prova Y do vestibular, muito antes deles chegarem ao terceiro ano do curso (Campos, 2014).

Para exemplificar mais possibilidades obtidas com uso de Big Data na Educação, Campos (2014) sugere que o uso de um *software* para acompanhar o rendimento individual e comparativo de diversos alunos de um mesmo curso, semestre após semestre, permite traçar correlações entre as notas obtidas e as características dos assuntos ensinados, como dificuldade, formato, tipo de interação, professor que ensinou o assunto, horário da aula etc. E, com isso, permite chegar a conclusões específicas e até individualizadas, como: “o

aluno X sempre obtém um bom rendimento quando um assunto, independentemente de seu grau de dificuldade, é ensinado pela manhã, através de vídeos e com professores recém formados”. Como resultado desta análise, os conteúdos são adaptados automaticamente ao perfil deste aluno. Isso significa que os sistemas educacionais baseados em Big Data são desenvolvidos para disparar relatórios de alertas automaticamente, sempre que alguma correlação interessante é calculada (Campos, 2014).

O *software*, alocado dentro de uma plataforma de *Internet*, e utilizável em qualquer computador, funciona como uma espécie de professor particular para o estudante, pois é capaz de identificar seus pontos fortes, aptidões e dificuldades e, em seguida, com base nesse diagnóstico, traçar planos de estudo, para aprimorar seu aprendizado. Quanto mais o usuário interage com a plataforma, mais eficiente ela vai se tornando, uma vez que reconhece melhor as necessidades de cada pessoa.

Porém, nem todo *software* de aprendizagem adaptativa é igual. A *Education Growth Advisors*, uma conceituada empresa de consultoria e assessoria do setor de Educação, divulgou um estudo para comparar os diferentes *softwares* desse tipo existentes no mundo. Para isso, criou um sistema próprio de classificação (Mayer-Schonberger e Cukier, 2015):

- As recomendações de conteúdos e caminhos são baseadas em um grupo de informações sobre o aluno, denominado perfil. A forma como esse perfil é criado varia de algoritmo para algoritmo. Alguns constroem esse perfil em uma etapa de diagnóstico do aluno e este perfil se mantém estático ao longo de todo seu percurso de aprendizado. Outros atualizam este perfil a cada interação do aluno com um conteúdo, com o *software* e com outros estudantes, aproveitando cada nova informação para aprender mais sobre o aluno e fazer recomendações melhores.
- A granularidade da personalização também varia. Em alguns casos, o algoritmo sugere um caminho entre cursos ou entre unidades e tópicos, sem especificar o material didático (ou objeto educacional) a ser estudado. Em outros casos, ele trata cada objeto educacional como um item e cada aluno recebe uma sequência personalizada de objetos - nesse caso, o aluno determina o que quer aprender e o algoritmo recomenda objetos que melhor o auxiliam para atingir esse aprendizado. O ideal é uma combinação das duas maneiras: determinar automaticamente as deficiências e,

a partir delas, recomendar objetos educacionais de maneira personalizada.

- O modelo de conteúdo é outro eixo de classificação das soluções adaptativas. Algumas delas já vêm com um conteúdo fixo. Outras permitem inserção de conteúdos da instituição de ensino mediante pedido. Há ainda as que oferecem ferramentas de autoria de conteúdos e permitem que cada instituição insira conteúdo diariamente.
- Os algoritmos encaram avaliações de maneiras diferentes. Para algumas soluções, as avaliações são pouco frequentes e, geralmente, usadas para diagnosticar um aluno e criar seu perfil. Outras vêm avaliações como parte do ensino e as personalizam para cada aluno.
- Há também uma grande variação na maneira como os *softwares* de aprendizado adaptativo se encaixam no contexto da instituição de ensino. Alguns podem ser usados somente em avaliações, em lições de casa etc. Outros, em todo o curso, inclusive no dia-a-dia de sala de aula. Acredita-se que a personalização em todo o curso é benéfica, mas entende-se que é melhor começar em partes do curso, para que as instituições de ensino consigam gerenciar as mudanças no desenho de sua sala de aula pouco a pouco.
- Por último, as soluções de aprendizado adaptativo suportam objetivos de aprendizagem diversos. Alguns deles, baseados na Taxonomia de Bloom³⁵, suportam o ensino de habilidades como, por exemplo, entender um tópico e lembrar-se dele. Outras auxiliam no ensino de habilidades de análise e aplicação de conhecimento. Há ainda a possibilidade de usar algoritmos de aprendizagem adaptativa para melhorar a capacidade de criação e avaliação dos alunos, mas nenhuma solução no mercado oferece esta terceira opção ainda. É impossível dissociar todos esses aspectos - entendimento, aplicação e criação - pois os três constituem elementos chave do processo educativo e não há, necessariamente, uma linearidade entre eles.

De acordo com Mayer-Schonberger e Cukier (2015), existem vários modelos além

³⁵A Taxonomia de Bloom é um instrumento para auxiliar a identificação e a declaração dos objetivos ligados ao desenvolvimento cognitivo que engloba a aquisição do conhecimento, competência e atitudes, visando facilitar o planejamento do processo de ensino e aprendizagem. (Fonte: <http://www.scielo.br/pdf/gp/v17n2/a15v17n2.pdf>)

dos que foram classificados pela *Education Growth Advisors* nos quais esses sistemas são baseados. Uns estão voltados para a armazenagem de informações sobre o material que está sendo ensinado, como a solução para uma questão ou aulas e tutoriais, e de metodologias de especialistas para ilustrar abordagens para as perguntas.

Outros são utilizados para determinar o nível de habilidade do aluno, através do método empregado no teste adaptativo informatizado (CAT, sigla em inglês para *Common Admission Test*), em que o assunto é apresentado com questões que são selecionados com base no nível de dificuldade e de habilidade presumida do assunto. Através de testes, o computador ajusta a pontuação de acordo com as respostas e seleciona perguntas com um grau maior de dificuldade à medida que as questões são respondidas corretamente. Para isso, um grande conjunto de perguntas é acumulado e classificado de acordo com a dificuldade, por meio da análise de peritos, experimentação, ou uma combinação dos dois. Em seguida, o computador executa uma busca binária a fim de dar ao assunto uma questão que está entre o que já foi estabelecido e confirmado como sendo os níveis máximos e mínimos de habilidade do usuário. Estes níveis são ajustados para o nível da dificuldade da questão e uma margem de erro é construída para permitir cenários em que a resposta do usuário não seja indicativo real de seu nível de habilidade, e sim apenas uma coincidência. Por exemplo, exibir várias perguntas de um mesmo nível de dificuldade reduz significativamente a probabilidade de uma resposta enganosa, e permite que o avanço para o próximo nível de habilidade assumido esteja livre de possíveis erros de avaliações.

Há também os algoritmos Richer³⁶ que são utilizados para lidar com a casualidade e fornecer um diagnóstico mais amplo dos pontos fracos do aluno. Ao invés de utilizar níveis simples de capacidade como no CAT, é realizada a ligação de conceitos às perguntas e são definidos pontos fortes e fracos em termos de conceitos.

Como vários conceitos podem influenciar uma única pergunta, estas têm de ser

³⁶Algoritmos Richer são utilizados para gerar redes aleatórias sem escala a partir de um mecanismo de fixação preferencial, o que significa que o nó mais conectado da rede é o mais provável de receber novas ligações. É similar ao que ocorre nas redes sociais, em que os nós fortemente ligados representam pessoas conhecidas com muitas relações: um link de A para B significa que a pessoa A “sabe” ou “está familiarizada com” a pessoa B. Quando um novo usuário entra na comunidade, um novo nó é adicionado à rede e ele tem mais chances de se familiarizar com as pessoas mais visíveis (nós mais conectados) do que com um parente desconhecido.

ligadas a todos os conceitos relevantes, por exemplo, através de uma matriz, que pode listar valores binários para a interseção de cada conceito e todas as perguntas. Em seguida, os valores de probabilidade condicional são calculados de modo a refletir a probabilidade de que um aluno que é fraco em um determinado conceito vai deixar de responder corretamente a uma questão específica. Essa probabilidade de fraqueza em todos os conceitos condicionais em respostas incorretas em todas as questões pode ser calculada baseada na Lei de Bayes³⁷, em que as probabilidades *a priori* são alteradas tendo em conta novas evidências de forma a obter probabilidades *a posteriori*. As lições são organizadas por conceito, ou seja, o plano de aula é adaptado em conformidade com a coleção de fraquezas após sua análise.

E ainda existem os modelos utilizados para incorporar as melhores ferramentas educacionais que a tecnologia tem a oferecer, como apresentações multimídia com conselhos de professor-especialista em métodos de apresentação. Esse modelo é conhecido como modelo de instrução, e nele, as aulas são classificadas em correspondência à fila de questões. Quando o nível do aluno é satisfatoriamente determinado, a lição apropriada é fornecida e quando as respostas incorretas estão sendo avaliadas, alguns sistemas fornecem *feedback* às questões reais em forma de dicas. Quando o aluno comete erros, sugestões úteis aparecem, como “analisar cuidadosamente o sinal do número”.

3.2 Aplicação de Big Data no Ensino

Cada pessoa aprende de uma maneira diferente. Por esse motivo, o conteúdo deve se adaptar ao aluno, e não o contrário. Existem plataformas que adaptam seu conteúdo de acordo com a necessidade de cada aluno, baseadas na velocidade de aprendizado sobre cada competência ou disciplina. Dessa forma, o curso oferecido nessas plataformas nunca é o mesmo para todos os alunos.

Este tópico visa apresentar alguns exemplos de plataformas que aplicam o Big Data em *sites* voltados para Educação, bem como suas especificidades. São eles: *Knewton*,

³⁷A Lei de Bayes é uma teoria da probabilidade que mostra a relação entre uma probabilidade condicional e a sua inversa, ou seja, a probabilidade de uma hipótese dada a observação de uma evidência e a probabilidade da evidência dada pela hipótese. Com essa lei é possível calcular a probabilidade de um evento A dado um evento B depende não apenas do relacionamento entre os eventos A e B, mas também da probabilidade marginal (ou “probabilidade simples”) da ocorrência de cada evento.

Khan Academy, Geekie Lab, Adaptativa, Enem Quiz, Ismart Online e Moodle.

3.2.1 Knewton

Lançada em 2008, a *Knewton*³⁸ é considerada a maior plataforma de Ensino Adaptativo do mundo e fornece a alunos, professores, pais e gestores públicos um diagnóstico preciso do aprendizado de cada aluno que usa a ferramenta.

Na plataforma, há aplicações para conteúdos didáticos que registram o seu uso pelos estudantes e coletam informações sobre seus desempenhos. Essas informações são processadas por algoritmos, baseados em milhões de pontos de dados, para adaptar aulas suplementares para qualquer estudante, em tempo real, e recomendar o que eles devem estudar. Dessa forma, os estudantes não só aprendem melhor, como mais rápido. Além disso, os professores podem personalizar lições e gerenciar classes com alunos que possuem ritmos diferentes de aprendizagem (Layton, 2015).

Inicialmente, o *software* disponibilizou conteúdos relacionados a lições de Matemática, Inglês, História e Ciências, todos submetidos à análise de uma equipe de professores e pesquisadores da Educação. No entanto, *Knewton* é uma plataforma aberta, ou seja, qualquer um pode fazer *upload* de aulas, conteúdos e avaliações. Isso significa que o conteúdo disponível irá crescer continuamente (Layton, 2015).

A fim de garantir a privacidade, os dados coletados na plataforma são criptografados e a *Knewton* afirma tomar todas as providências para assegurar que as informações estão seguras, inacessíveis e não serão vendidas ou compartilhadas.

3.2.2 Khan Academy

A *Khan Academy*³⁹ é uma ONG educacional cuja missão é fornecer educação de alta qualidade para todos em qualquer lugar.

“A *Khan Academy* oferece exercícios, vídeos de instrução e um painel de aprendizado personalizado que habilita os estudantes a aprender no seu próprio ritmo dentro e fora da sala de aula. ”

A plataforma surgiu em 2008, quando seu fundador, Salman Khan, passou a

³⁸Knewton. Fonte: <https://www.knewton.com/>

³⁹Khan Academy. Fonte: <https://pt.khanacademy.org/>

postar vídeos no *Youtube* para ajudar uma prima a estudar Matemática. Atualmente, possui uma equipe diversificada de desenvolvedores, professores, *designers*, estrategistas, cientistas e especialistas em conteúdo, que se reuniu para levar educação de alta qualidade para todos em qualquer lugar.

Na plataforma, são abordados assuntos de áreas diversas, como: Matemática, Ciência, Programação de computadores, História, História da arte, Economia etc., no entanto, apenas as lições de Matemática são ensinadas de forma adaptativa.

“[...] as lições de matemática guiam os estudantes do jardim de infância até o cálculo, usando tecnologias adaptativas de ponta que identificam os pontos fortes e lacunas no aprendizado. ”

Quando o aluno utiliza a plataforma pela primeira vez, ele é submetido a um teste com oito questões para que o *software* conheça quem é o usuário e determine quais conhecimentos ele já sabe. A partir disso, ele já sugere uma série de conteúdos.

Para garantir que o aluno realmente aprendeu determinado conteúdo, a plataforma faz uma verificação periódica dos assuntos que o aluno acertou. Com a ajuda de um algoritmo, o programa pergunta os conhecimentos diversas vezes no formato de desafios. Se o aluno errar, o programa se adapta para ensinar o conteúdo ao aluno novamente.

A ferramenta também permite cadastrar outras pessoas para visualizar os resultados de um estudante - professores ou tutores que auxiliem o aluno nos estudos - e fornece um resumo de desempenho da turma como um todo, além de perfis detalhados dos alunos. Com isso, a organização busca habilitar os tutores de todos os tipos a compreender melhor o que seus filhos ou alunos estão fazendo, e a melhor forma de ajudá-los no processo de aprendizado.

“A plataforma é capaz de identificar por quanto tempo os vídeos foram assistidos, quantas tentativas foram necessárias para a resolução de um exercício e quanto tempo os alunos ficaram conectados à plataforma. [...] o *Khan Academy* vem trabalhando em um modelo de previsão de comportamento, no qual será possível avaliar mais precisamente os ganhos de aprendizagem e que peças de conteúdo devem ser sugeridas individualmente para cada estudante. Essa escala de análise tem sido capaz de gerar mais significado para os dados, enquanto alunos e professores utilizam o *Khan Academy*. ” (Scaico et al, 2014)

Além de ter parcerias com instituições renomadas como a NASA, o Museu de Arte Moderna, a Academia de Ciências da Califórnia e o MIT para oferecer conteúdo

especializado, a *Khan Academy* fez parceria com a Fundação Lemann para traduzir os exercícios da plataforma para o português.

3.2.3 Geekie Lab

Pioneira na promoção de Ensino Adaptativo no Brasil, a plataforma *Geekie*⁴⁰ foi fundada em 2011 “com o objetivo de melhorar o desempenho dos estudantes da educação básica, das redes pública e privada do Brasil e, assim, aumentar suas chances de ingressar no ensino superior”.

“O *software*, alocado dentro de uma plataforma de *Internet*, e utilizável em qualquer computador, funciona como uma espécie de professor particular para o estudante, pois é capaz de identificar seus pontos fortes, aptidões e dificuldades e, em seguida, com base nesse diagnóstico, traçar planos de estudo, para aprimorar seu aprendizado. Quanto mais o usuário interage com a plataforma, mais eficiente ela vai se tornando, uma vez que reconhece melhor as necessidades de cada pessoa.”

Além de mapear o aprendizado dos alunos, o serviço também permite avaliar o desempenho dos professores:

“Se mais de uma turma do mesmo professor apresenta dificuldade no mesmo ponto, significa que o conteúdo não foi passado ou, então, que foi passado de forma que os alunos não conseguiram assimilar.”

A *Geekie* atua com base no conceito “*one pay, one free*”, ou seja, para cada aluno de escola particular com acesso ao *software*, um estudante da rede pública também ganha o benefício gratuitamente, por meio de parcerias com as Secretarias Estaduais de Educação.

Criada com o intuito de ajudar na preparação dos alunos para o Exame Nacional do Ensino Médio (Enem), colaborando para o aprendizado, estima-se que a ferramenta produziu 3,7 milhões de planos de estudo personalizados e elevou em até 30% o desempenho de seus usuários, sendo que 70% dos alunos inicialmente diagnosticados como “fracos” se tornaram “proficientes”.

A plataforma disponibiliza simulados nas quatro áreas do exame, baseados na Teoria de Resposta ao Item (TRI), método estatístico usado pelo Enem para avaliar os candidatos em seus diferentes níveis de habilidades e competências, o que aproxima o desempenho no simulado realizado na plataforma da nota real da prova.

⁴⁰Geekie. Fonte: <http://redepapelsolidario.org.br/plataforma-geekie-e-pioneira-na-promocao-de-ensino-adaptativo-brasil/>

A *Geekie* é considerada pelo Ministério da Educação a única organização certificada pelo Inep (Instituto Nacional de Estudos e Pesquisas Educacionais) como tecnologia confiável de simulação do Enem.

3.2.4 Adaptativa

Assim como o *Geekie*, a Adaptativa⁴¹ também oferece vídeos com explicações de conteúdos e listas de exercícios em um plano de estudos, que são uma combinação das provas do Enem, cerca de cinco anos dos vestibulares da Fuvest, Unicamp e Unesp, e 5 mil exercícios de fixação desenvolvidos pelos professores do programa, todos resolvidos em vídeo.

A plataforma disponibiliza vídeo-aulas que cobrem de maneira simples e didática todas as áreas avaliadas no Enem e simulados *online* que contém questões inéditas e são corrigidos com a mesma metodologia do Enem, o que permite determinar como será o desempenho do aluno na prova real.

A Adaptativa é atualizada todo mês com 200 questões inéditas e fornece relatórios de diagnósticos detalhados que mostram o progresso do aluno e os tópicos nos quais ele precisa focar para ter um bom desempenho no Enem. Além disso, a plataforma conta com um método moderno de avaliação de redação *online*, único no Brasil, em que a correção é colaborativa, ou seja, o aluno escreve e participa da correção, aprendendo a escrever na prática.

As questões resolvidas pelos estudantes são classificadas por assunto, nível de dificuldade e pelas competências do Enem. Ao finalizar um exercício, o aluno recebe a porcentagem de seus acertos de questão e das habilidades do Enem e pode comparar com a média dos outros usuários.

Segundo dados da plataforma, entre os conteúdos que mais são cobrados no Enem estão: História do Brasil independente e História contemporânea; Globalização da economia, Comunicações e transportes em Geografia; Ecologia e Fisiologia humana em Biologia; Funções orgânicas e inorgânicas em Química; Visualização de dados em Matemática; Cinemática e energia em Física e Interpretação de textos e literatura pós-moderna em Português. Os assuntos que os alunos mais erram nos exercícios envolvem questões in-

⁴¹ Adaptativa. Fonte: www.adaptativa.com.br

terdisciplinares, que combinam conteúdos de múltiplas disciplinas e questões teóricas de ciências naturais, que requerem a aplicação prática da teoria.

3.2.5 Enem Quiz

O Enem Quiz⁴² é uma ferramenta voltada aos candidatos que se preparam para o Exame Nacional do Ensino Medio (Enem), que oferece uma preparação *online* com a revisão das matérias exigidas no edital e com questões de edições anteriores da prova. A ferramenta dá acesso a um número ilimitado de questões, um simulador de diferentes disciplinas e assuntos, além de fornecer uma estatística completa de desempenho e acesso aos comentários dos professores sobre as questões resolvidas e vídeo-aulas dos conteúdos exigidos no Enem.

“O Enem Quiz é um *site* gratuito que oferece aulas e simulados para estudantes que farão o Enem. Nele, os estudantes podem fazer simulados, assistir aulas de diferentes disciplinas, além de contar com materiais de apoio e acompanhar o desempenho com dados sobre seus estudos. ”

Na plataforma, o aluno encontra vídeos de 10 minutos, com explicações dos conteúdos que caem no Enem. A cada aula assistida, um simulado é gerado com questões relacionadas. São cerca de 3 mil questões, envolvendo todas as edições anteriores do Enem e questões próprias da plataforma, criadas por uma equipe de 15 professores, todas resolvidas em vídeo. Portanto, se o aluno errou a questão, pode assistir ao vídeo no mesmo momento e descobrir qual foi o erro.

Além disso, para questões que envolvem mais de uma disciplina, o programa disponibiliza vídeos em um canal no *YouTube* com as resoluções comentadas pelos professores de todos os conteúdos, relacionando uns com os outros. A plataforma não possibilita a interação direta com o professor.

Há também versões para dispositivos móveis (Android e iOS) que permitem a avaliação de conhecimento através de testes aplicados nas outras edições do Enem. Nesses aplicativos, é possível criar um simulado personalizado com as matérias e o número de questões desejados, e definir de quais anos selecionar as perguntas. O Enem Quiz mostra o tempo decorrido e também em qual pergunta o usuário está, além de permitir marcar questões como favoritas. Ao final da prova, o aplicativo grava as estatísticas de tempo e

⁴²Enem Quiz. Fonte: <http://enemquiz.com.br/>

conteúdo avaliados. No iOS (iPhone e iPad), o Enem Quiz é ainda mais completo, com vídeo aulas e comentários em vídeo das questões.

3.2.6 Ismart Online

O Instituto Social para Motivar, Apoiar e Reconhecer Talentos (*Ismart*)⁴³ foi criado em 1999 e “é uma entidade privada, sem fins lucrativos, que identifica jovens talentos de baixa renda, de 14 a 15 anos de idade, e lhes concede bolsas em escolas particulares de excelência e o acesso a programas de desenvolvimento e orientação profissional, do ensino fundamental à universidade e busca apoiar o desenvolvimento de jovens talentos acadêmicos”.

O instituto tem 3 projetos de bolsas: Alicerce voltado para alunos do Ensino Fundamental; Bolsa Talento para alunos que estão ingressando no Ensino Médio; e Bolsa Universitários para alunos que querem cursar faculdade.

Além desses projetos, há o *Ismart Online*, criado para dar mais oportunidades de desenvolvimento a jovens talentos e prepará-los para os desafios do ensino médio.

A plataforma de ensino *online* é dividida em dois módulos⁴⁴:

- De um lado, os alunos acessam conteúdos de Português e Matemática em uma plataforma adaptativa desenvolvida pela *Geekie*, *startup* de tecnologia aplicada à educação, parceira do Ismart na iniciativa. Após a realização de simulados, a plataforma faz um diagnóstico preciso e específico para determinar quais são os pontos fortes e pontos de melhoria de cada usuário.
- No Módulo de Cultura, o Ismart publica conteúdos com uma linguagem leve e divertida, mas com mensagens importantes para o desenvolvimento pessoal dos alunos. O objetivo é trabalhar aspectos ligados a autonomia, inspiração, motivação e persistência. Como parte do módulo, são realizados encontros presenciais para promover a integração da comunidade. Nesses *workshops*, os alunos desenvolvem projetos, aprendem a trabalhar em equipe e colocam em prática os conhecimentos trabalhados virtualmente. Em grupos, os jovens também resolvem lições - para

⁴³Ismart. Fonte: <http://www.ismart.org.br/quem-somos/>

⁴⁴Ismart - módulos. Fonte: <http://www.ismart.org.br/2015/11/ismart-convoca-429-alunos-para-o-projeto-ismart-online-em-2016/>

isso, contam com o auxílio de mentores, que são universitários e recém-formados da comunidade de bolsistas do *Ismart* e da Fundação Estudar.

Além de manter a bolsa, quem se destaca no projeto recebe prêmios como *tablets* e vales-presente.

3.2.7 Moodle

A Plataforma Moodle - *Modular Object Oriented Distance Learning* (Objeto Modular Orientado ao Ensino a Distância) - também pode ser entendida como um Ensino Adaptativo, uma vez que ela incorpora ferramentas básicas de adaptação da aprendizagem, e visto que o autor/tutor de um determinado curso pode requerer determinados recursos e atividades. O autor/tutor pode ainda empregar um instrumento de avaliação para reconhecer o estilo e o resultado do que foi aprendido pelo aluno (Lima, 2012).

Outro exemplo de que o Moodle é uma plataforma adaptativa pode ser observado no módulo Questionário⁴⁵, quando a opção “Modo Adaptativo” é selecionada para o campo “Como se comportam as questões”. Esse campo determina como os alunos interagem com as perguntas do questionário.

Quando o Modo Adaptativo é escolhido para uma questão, o aluno pode tentar acertá-la várias vezes antes de passar para a questão seguinte. Se a resposta for incorreta, ele poderá tentar novamente imediatamente. No entanto, ele poderá ser penalizado (geralmente, sua pontuação é subtraída para cada tentativa errada). Além disso, a questão pode se adaptar de acordo com a resposta do aluno, sendo possível convidá-lo a tentar novamente, dar dicas antes que ele faça uma nova tentativa e até mesmo aumentar o grau de dificuldade da pergunta.

“Adaptação permite aos autores criar itens para uso em situações formais que ajudam a guiar os candidatos através de uma dada tarefa e, ao mesmo tempo, proporcionar um resultado que leva em consideração o caminho seguido. ”

⁴⁵Moodle - Questionário. Fonte: <http://www2.tjam.jus.br/moodle/help.php?module=quizfile=adaptive.html>

4 Análises

Assim como as demais áreas de atuação da sociedade, a Educação tem utilizado cada vez mais as TIC (Tecnologias de Informação e Comunicação) para apoiar o processo de aprendizagem, tornando-o cada vez mais ágil e fazendo com que os estudantes estejam cada vez mais bem preparados para o mundo digital.

Nesse cenário, surgiram os Ambientes Virtuais de Aprendizagem (AVA) que, segundo Ribeiro et al (2007)

“são *softwares* educacionais via *Internet*, destinados a apoiar as atividades de educação a distância. Estes *softwares* oferecem um conjunto de tecnologias de informação e comunicação, que permitem desenvolver as atividades no tempo, espaço e ritmo de cada participante”.

Os Ambientes Virtuais de Aprendizagem, também conhecidos como plataformas de EAD, podem ser utilizados em: atividades presenciais, possibilitando aumentar as interações para além da sala de aula; em atividades semipresenciais, nos encontros presenciais e nas atividades à distância; oferecendo suporte para a comunicação e troca de informações e interação entre os participantes (Ribeiro et al, 2007). Eles já fazem parte do dia a dia do processo de ensino, e quando utilizados pelos estudantes, deixam rastros e estatísticas *online*, o que gera uma grande quantidade de dados educacionais que podem fornecer às instituições de ensino conhecimentos importantes sobre o processo de ensinar e aprender. Isso permite que elas melhorem seus modelos de gestão de aprendizagem e de avaliação, construam novas práticas de ensino, discutam mais as políticas educacionais existentes e promovam novas experiências de aprendizagem capazes de melhorar a maneira como os professores ensinam, os estudantes aprendem e a escola funciona (Scaico et al, 2014).

No entanto, ferramentas tradicionais não são capazes de coletar, processar e analisar essa grande quantidade de dados de forma a apontar evidências, padrões e tendências que podem aprimorar e inovar estratégias de ensino e aprendizagem. Isso só é possível com a utilização de técnicas relacionadas ao Big Data, devido ao seu alto poder de processamento e análise.

O Big Data está relacionado à capacidade de analisar qualquer tipo de informação digital (em grande volume e variedade), em tempo real, sendo fundamental para a tomada de decisões. O conceito considera não somente grandes quantidades de dados, a velocidade de análise e a disponibilização destes, como também a relação com e entre os volumes (Alecrim, 2015).

No campo educacional, a utilização de Big Data denomina-se Ensino Adaptativo e está permitindo capturar muitas informações sobre os estudantes e sobre sua interação com os conteúdos instrucionais, ambientes de aprendizagem e a forma como são avaliados. Com isso, os processos de avaliação, *feedback* e entrega de conteúdo estão se tornando mais eficazes (Scaico et al, 2014).

De acordo com a pesquisa *NMC Horizon Report: 2014 Higher Education Edition*, o Big Data é considerado uma área de pesquisa de grande relevância para educação do futuro, devido a seu grande potencial para o suporte à construção de ambientes adaptativos de aprendizagem, capazes de mapear o comportamento e os interesses dos estudantes e os fatores que podem levá-los a níveis maiores de engajamento (Scaico et al, 2014). Além disso, pode “conduzir mais eficientemente o planejamento de cursos, apontando como um conteúdo pode ser melhor dividido em tópicos, como estes devem ser sequenciados e que configurações podem gerar curvas maiores de aprendizagem” (Scaico et al, 2014).

Prover tecnologia para as instituições de ensino é uma das melhores formas de auxílio ao educador que, atualmente, busca se transformar em um mediador entre os alunos e o processo de construção do conhecimento. No Ensino Adaptativo, os alunos são vistos como criadores de conhecimento e não mais como consumidores como nos modelos educacionais tradicionais. Por isso, é preciso que sejam realizadas mudanças pedagógicas para estimular a criatividade dos estudantes e o interesse dos mesmos pelo aprendizado.

Porém, a tecnologia precisa colaborar para o crescimento da qualidade de ensino sem que as raízes que dão base à tecnologia educacional sejam perdidas, pois o uso desta só será eficaz se for agregado ao conhecimento e ao desenvolvimento dos docentes. Isso significa que o Ensino Adaptativo atua apenas como um complemento dos esforços educacionais e não substituirá o professor, a quem sempre caberá o papel final da avaliação sistêmica da aprendizagem do aluno.

“Apesar do processamento de grandes volumes de dados ser praticamente todo automatizado, a participação dos seres humanos é fundamental. Isto ocorre, por exemplo, no processo de análise visual dos dados, que ajuda a criar mais sentido para os mesmos ou na detecção inicial de padrões. ” (Scaico et al, 2014)

O uso das técnicas de Big Data na Educação serve para potencializar alternativas no desenvolvimento dos processos de ensino e de aprendizagem, mas exige uma evolução didática do professor, que precisa saber explorar as potencialidades das ferramentas que estão ao seu alcance e fazer uso delas para mediar e instigar no aluno a busca pelo conhecimento. Nesse sentido, o papel do professor não se limita à transmissão de conhecimento, ele assume diferentes perfis, como: validador, curador, avaliador, pesquisador, conselheiro, mediador, colaborador, neuroeducador e facilitador.

Não basta gerar e/ou captar um grande volume de dados dos estudantes, é preciso conseguir e saber usá-los. O Ensino Adaptativo pode auxiliar a atuação de educadores, gestores e formuladores de políticas educacionais e, conseqüentemente, favorecer melhorias nos ganhos de aprendizagem. Entretanto, Big Data é um conceito relativamente novo, que precisa ser compreendido pelos estudiosos e interessados na área de Educação para que novas soluções que utilizam a tecnologia possam ser projetadas. Para isso, é preciso conhecer as tendências e as tecnologias que estão surgindo nesse cenário.

Em termos gerais, as ferramentas de análise de Big Data permitem o tratamento de grandes quantidades de informações, sejam elas estruturadas ou não, para gerar novas análises e entendimento sobre as informações existentes. Essas ferramentas vão além do banco de dados relacional e permitem correlacionar dados de diversas fontes e nos mais variados formatos, como documentos, planilhas, *e-mails*, imagens, áudios e até vídeos (Ulisses, 2015).

Como o foco do Big Data é apenas o dado, ele permite executar consultas que seriam inviáveis em termos de tempo de execução e recursos computacionais nas soluções tradicionais, pois estas consideram a indexação, o relacionamento e a validação dos dados em suas consultas. Contudo, isso não determina o fim do SQL (*Structured Query Language*), comumente utilizado em SGDB (Sistema de Gerenciamento de Banco de Dados), pois ferramentas de desenvolvimento para Big Data, como *Cassandra*, *MongoDB* e *Hadoop*, foram desenvolvidas de forma a suportar a linguagem SQL. Além disso, os bancos

de dados relacionais continuarão por muito tempo úteis a uma série de aplicações (Ulisses, 2015).

A necessidade de armazenar e tratar dados não estruturados levou a criação dos bancos de dados NoSQL, onde “No” significa “*Not Only*”. O NoSQL faz referência às soluções de bancos de dados que possibilitam o armazenamento de diversas formas, não se limitando ao modelo relacional tradicional. Bancos desse tipo são mais flexíveis, e compatíveis com as premissas BASE (*Basically Available, Soft state, Eventually consistency* - Basicamente disponível, Estado Leve, Eventualmente consistente) que competem com as premissas mais restritivas ACID (*Atomicity, Consistency, Isolation e Durability* - Atomicidade, Consistência, Isolamento e Durabilidade) dos bancos relacionais (Alecrim, 2015).

Escalar um banco de dados NoSQL é mais fácil e menos custoso. Isso porque ele conta com propriedades mais flexíveis e são otimizados para trabalhar com processamento paralelo, distribuição global (vários *data centers*) e aumento imediato de sua capacidade. Há várias categorias de bancos de dados NoSQL, por isso, as soluções do tipo atendem à grande variedade de dados que existe, tanto estruturados, quanto não estruturados, como: bancos de dados orientados a documentos, bancos de dados chave/valor, bancos de dados de grafos etc (Alecrim, 2015).

Nos tempos atuais, os avanços computacionais nos permitem guardar, organizar e analisar dados com mais facilidade e com uma frequência muito maior. O Big Data surge como alternativa para lidar com essa “explosão” de dados, uma vez que as ferramentas computacionais convencionais já não conseguem fazê-lo satisfatoriamente. Devido à grande quantidade de dados gerada e armazenada, estruturas centralizadas de processamento estão sendo substituídas por soluções distribuídas e com elasticidade, ou seja, mais flexíveis e capazes de suportar aplicações com volumes de dados que crescem substancialmente em pouco tempo. Por isso, as aplicações devem ser capazes de consumir pequenas partes das informações e gerar pequenas partes de dados processados, que serão calculados em conjunto para criar o resultado final. Os resultados geralmente são arquivos que contém resumos das informações, agrupadas conforme necessário, de forma a possibilitar o entendimento e a tomada de decisão sobre aqueles dados (Ulisses, 2015).

Portanto, “questões de padronização das informações e da interoperabilidade dos sistemas que processarão os dados ainda precisam ser discutidas para que se possa garantir a escalabilidade do conhecimento que está sendo produzido” (Scaico et al, 2014).

Apesar de todo benefício que o Big Data pode oferecer, é preciso ter cautela quanto à adoção dessas ferramentas, pois, geralmente, as consultas são executadas em ambiente distribuído e isso faz com que o tempo de execução de pequenas consultas seja muito maior do que em um SGDB comum. Dessa forma, essas ferramentas são indicadas para trabalhar com grandes quantidades de dados, onde o *overhead* da computação distribuída é insignificante frente à quantidade de dados a serem processados.

Ademais, os custos de produção de ferramentas robustas que sejam capazes de lidar com o processamento dos dados nestas dimensões ainda são muito altos, assim como a formação de especialistas para trabalhar em uma área que exige a compreensão de conceitos multidisciplinares, como Computação, Estatística, Psicologia, dentre outras. O profissional de Big Data é um cientista de dados que exerce papel fundamental no processo de conversão de *bytes* em vantagem competitiva às organizações. Geralmente, essa função demanda formação em Ciência da Computação e Matemática, e habilidades analíticas necessárias para encontrar a providencial agulha no palheiro de dados recolhidos pela empresa. Mas isso não é regra, pois como afirmaram Mayer-Schonberger e Cukier (2013), os cientistas de dados devem ser profissionais que conhecem profundamente o negócio e tenham imaginação e criatividade para fazer as perguntas certas, e não necessariamente serão encontrados no setor de TI.

Os autores também sinalizam para “a necessidade de profissionais especializados para trabalhar como revisores das análises e previsões do Big Data, evitando, assim, formas mais invasivas de regulação” (Mayer-Schonberger e Cukier, 2013).

Taurion (2013) afirma que o grande desafio do Big Data nos próximos anos será ter profissionais capacitados a realizarem análises preditivas que demandam capacitação em Estatística, Matemática e conhecimento das regras de negócio. Esses profissionais também precisam estar familiarizados com a modelagem de dados não estruturados.

Taurion (2013) identifica três perfis básicos de profissionais engajados em Big Data:

- a) Cientistas de dados, como descrito acima. Profissionais capacitados em Estatística, Ciência da computação e/ou Matemática capazes de analisar grandes volumes de dados e extrair deles *insights* que criem novas oportunidades de negócio;
- b) Analistas de negócio que conhecendo bem o negócio em que atuam consigam formular as perguntas corretas, analisar as respostas e tomar decisões estratégicas e táticas que alavanquem novos negócios ou aumentem a lucratividade da empresa. Esta função tende a ser acoplada às funções do cientista de dados;
- c) Profissionais de tecnologia que cuidarão da infraestrutura e seu suporte técnico para suportar Big Data. O aparato tecnológico de Big Data não é muito comum em empresas tipicamente comerciais, pois demanda *expertise* em gerenciar *hardware* em *clusters* de alta performance (*Hadoop* é massivamente paralelo) e pensar em volumes de dados significativamente maiores e muito mais variados que comumente se usa em sistemas tradicionais.

A escassez de profissionais capacitados a trabalhar com Big Data é um desafio, pois “abre muitas perspectivas profissionais para os que abraçarem a função, mas também atuará como um entrave, pois dificultará às empresas usarem Big Data com eficiência” (Taurion, 2013).

A relação entre Big Data e privacidade de dados não está bem definida e é uma questão que precisa ser discutida. A imensa quantidade de dados gerados na *Web* evidencia a viabilidade ética de replicar essas informações ou usá-las para controle governamental ou fins empresariais.

De acordo com o que foi divulgado pelo jornal The Washington Post, a Agência Nacional de Segurança dos Estados Unidos e o FBI têm acesso direto aos servidores centrais das mais importantes empresas de *Internet* do planeta, como *Google*, *Facebook*, *Skype* e *Aol*. Dados oriundos dessas empresas, como áudios, vídeos, fotografias, conteúdos de *e-mails*, arquivos transferidos e conexões dos usuários, são totalmente monitorados pelas agências de inteligência norte-americanas através do programa Prisma⁴⁶, o que evidencia

⁴⁶Prisma. Fonte: <http://veja.abril.com.br/noticia/mundo/eua-governo-vasculha-dados-de-fontes-como-google-e-facebook/>

a inviabilidade de manter informações absolutamente seguras na rede e a necessidade de impor limites éticos à captura de dados para geração de informações.

Em 2013, no Brasil essa questão também veio à tona quando o Serasa firmou um acordo com o Tribunal Superior Eleitoral (TSE) para ter acesso aos dados de eleitores brasileiros, trato que foi desfeito após o veto da presidente do TSE na época. Atualmente, também está em discussão no país a introdução da primeira lei de proteção de dados pessoais, que tem o intuito de dar ao cidadão o controle sobre suas informações pessoais, listando os direitos sobre seus próprios dados e determinando formas adequadas de tratamento das informações por entidades do setor público ou privado (Rocha, 2015).

“Nota-se, entretanto, a desinformação dos usuários das redes sociais sobre os possíveis usos das informações disponibilizadas nestes sítios. Surge a questão ética no uso desses dados que precisa ser debatida com maior profundidade com o objetivo de garantir o respeito ao consumidor. É necessário que o indivíduo tenha maior controle sobre quem pode acessar suas informações e qual o uso que as empresas estão fazendo destes dados, porém o tema ainda carece de regulamentações que estabeleçam padrões de identificação e de segurança dos dados pessoais. ” (Novo et al, 2014)

Taurion (2013) afirma que simples alterações das regras vigentes não serão suficientes para garantir a privacidade dos dados e que as mudanças devem ser práticas e eficientes. Para o autor, é possível que o conceito de privacidade implique em uma maior responsabilização dos usuários dos dados por seus atos, uma vez que eles precisam ter tanto o ônus da análise, quanto a responsabilidade pelo uso das informações.

No campo educacional, também é preciso preservar a privacidade dos dados de estudantes e professores.

“[...] a extração de conhecimento que ocorre com a análise das bases de dados e com a tentativa de compreender determinados comportamentos, assim como própria onipresença da tecnologia no ambiente educacional, têm levantado a uma discussão sobre inúmeras questões éticas. A extração e a manipulação de dados impõem a necessidade de regulação de ações para que se possa manter os princípios da privacidade, confidencialidade e transparência. ” (Scaico et al, 2014)

A conclusão a que este trabalho chega é a de que o Big Data reflete um cenário real, pois cada vez mais grandes volumes de dados são gerados e isso exige uma abordagem capaz de aproveitá-los ao máximo e, assim, “tirar proveito de suas análises para agir mais acertadamente, experimentando com velocidade e explorando mais pistas sem que, para

isso, seja necessária uma compreensão mais detalhada dos fatos” (Mayer-Schonberger e Cukier, 2013).

O Big Data se posiciona como um modelo evolutivo da análise de dados tradicionais, suprimindo novas demandas de análise rápida dos dados advindos de diversas fontes e em maior quantidade. Está claro o crescimento pelo interesse e estudo ao redor do tema embora este ainda seja maior na esfera empresarial do que acadêmica. (Novo et al, 2014)

5 Considerações finais

As técnicas de análise de dados do Big Data serão indispensáveis para a solução dos problemas mais urgentes, como o enfrentamento de mudanças climáticas, a erradicação de doenças, a promoção da boa governança e o desenvolvimento econômico. No campo de Recursos Humanos, por exemplo, o mérito da análise de dados é fornecer as evidências para apoiar as decisões tomadas por gestores de pessoas; na área comercial, o Big Data permite entender o perfil do consumidor e o seu comportamento nas redes sociais, bem como o hábito de consumo, intenção de compras, relevância e influência; na Economia, o Big Data está permitindo que novas fontes de dados sejam utilizadas para definir e monitorar comportamentos econômicos, uma vez que os resultados mostram que os indicadores econômicos de custo-benefício podem ser construídos com base em conjuntos de dados publicamente disponíveis de mídia social; e na Educação, o uso de Big Data permite gerar conhecimentos generalizáveis sobre o processo de ensino-aprendizagem e as dificuldades inerentes a determinados conteúdos ou estratégias pedagógicas, contribuindo com mudanças criativas e significativas para este campo.

Com o desenvolvimento deste trabalho foi possível identificar que o Big Data é considerado uma área de pesquisa de grande relevância para Educação do futuro, devido ao seu grande potencial para o suporte à construção de ambientes adaptativos de aprendizagem, capazes de mapear o comportamento e os interesses dos estudantes e os fatores que podem levá-los a níveis maiores de engajamento. Além disso, o Big Data permite conduzir o planejamento de cursos, apontando como um conteúdo pode ser melhor dividido em tópicos, em que sequência e que configurações podem gerar curvas maiores de aprendizagem. Com o Big Data, é possível tratar grandes quantidades de dados, estruturados ou não, para gerar novas análises e entendimento sobre as informações existentes e apresentar uma nova forma de planejar e fazer Educação através do Ensino Adaptativo.

Este trabalho apresenta as principais definições do termo Big Data, bem como os conceitos e variáveis que o fundamentam: os 5Vs - volume, variedade, velocidade, veracidade e valor. Além disso, cita exemplos de sua utilização nas mais variadas áreas

de atuação da sociedade. Todavia, enfatiza como essa tecnologia tem sido aplicada à Educação para ilustrar como ocorre o processo de Ensino Adaptativo, com o intuito de auxiliar as escolas na escolha de ferramentas de Big Data para a inclusão da metodologia de Ensino Adaptativo em suas instituições.

A pretensão deste trabalho é dar início a futuras discussões sobre o assunto para que os estudiosos e interessados na área de Educação possam pensar em novas soluções que utilizam essa tecnologia a fim de aprimorar o processo de ensino-aprendizagem e propor que trabalhos futuros façam a comparação das ferramentas aqui apresentadas, enumerando que tecnologias cada uma usa para se transformar em adaptativa ou quais técnicas são utilizadas nessas ferramentas para conhecer o aluno e avaliar a aprendizagem.

Referências Bibliográficas

- Alecrim, E. **O que é big data?** <http://www.infowester.com/big-data.php>, mar 2015. Último acesso: 14/02/2016.
- Marr, B. **Big data explained in less than 2 minutes - to absolutely anyone.** <https://www.linkedin.com/pulse/big-data-explained-less-than-2-minutes-absolutely-anyone-bernard-marr?trk=mp-reader-card>, mar 2015. Último acesso: 30/01/2016.
- Barros, H. **Big data: evolução ou re-evolução?** <http://www.administradores.com.br/artigos/tecnologia/big-data-evolucao-ou-re-evolucao/92126/>, dez 2015. Último acesso: 30/01/2016.
- Bettocchi, E. **Ensino adaptativo**, ago 2015.
- Anderson, C. **The end of theory: The data deluge makes the scientific model obsolete.** <http://www.wired.com/2008/06/pb-theory/>, 2008. Último acesso: 24/01/2016.
- Campos, N. M. **Ensino adaptativo: O big data na educação.** <http://educacao.estadao.com.br/blogs/a-educacao-no-seculo-21/ensino-adaptativo-o-big-data-na-educacao/>, abr 2014. Último acesso: 15/02/2016.
- Canary, V. P. **A tomada de decisão no contexto do big data**, 2013.
- Correia, M. B. F. **A comunicação de dados estatísticos por intermédio de infográficos: Uma abordagem ergonômica**, 2009.
- Michael Cox, D. E. **Application-controlled demand paging for out-of-core visualization.** <https://www.nas.nasa.gov/assets/pdf/techreports/1997/nas-97-010.pdf>, 1997.
- Cukier, K. **Tdata, data everywhere.** <http://www.economist.com/node/15557443>, 2010. Último acesso: 24/01/2016.
- e L. Prusak, T. H. D. **Knowledge: how organizations manage what they know**, 1998.
- Davenport, T. H. **Big data no trabalho**, 2014.
- Dumbill, E. **Big data now: 2012 edition**, 2012.
- Faria, L. **Big data na educação: como é usado nas escolas ao redor do mundo.** <http://www.datastorm.com.br/como-o-big-data-esta-sendo-usado-pelas-escolas-ao-redor-do-mundo/>, mar 2015. Último acesso: 14/02/2016.
- A. B. H. Ferreira, M. Anjos, M. B. F. **Novo aurélio, século xxi: o dicionário da língua portuguesa**, 1999. p. 2128.

- Ferreira, C. **Big data e a educação.** <http://www.gazetadopovo.com.br/blogs/educacao-e-midia/big-data-e-a-educacao/>, jul 2014. Último acesso: 14/02/2016.
- Frangomeni, A. H. **Dicionário enciclopédico de informática**, 1986.
- Gil, A. C. **Como elaborar projetos de pesquisa**, 2008. 5. edicao.
- Laney, D. **3d data management: Controlling data volume, velocity and variety**, 2001.
- Layton, L. **A personal tutor, available at the click of a mouse.** https://www.washingtonpost.com/local/education/a-personal-tutor-available-at-the-click-of-a-mouse/2015/08/26/ed05ac70-4aa0-11e5-8e7d-9c033e6745d8_story.html, ago 2015. Último acesso: 15/02/2016.
- de Sousa Lima, L. **Ensino adaptativo.** <https://www.portaleducacao.com.br/pedagogia/artigos/14870/ensino-adaptativo>, jul 2012. Último acesso: 15/02/2016.
- Machado, J. L. A. **Big data na educação.** <http://www.planetaeducacao.com.br/portal/artigo.asp?artigo=2591>, jul 2015. Último acesso: 15/02/2016.
- Castells, M. **A galáxia da internet: Reflexões sobre a internet, os negócios e a sociedade**, 2003.
- James Manyika, Michael Chui, B. B. **The next frontier for innovation, competition, and productivity.** http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation, 2011. Último acesso: 24/01/2016.
- Marr, B. **A brief history of big data everyone should read.** <https://www.linkedin.com/pulse/brief-history-big-data-everyone-should-read-bernard-marr>, fev 2015. Último acesso: 23/01/2016.
- Viktor Mayer-Schonberger, K. C. **Big data: Como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana**, 2013.
- Viktor Mayer-Schonberger, K. C. **Learning with big data**, 2014.
- Miyazaki, M. **A brief history of data analysis.** <https://www.flydata.com/a-brief-history-of-data-analysis>, 2015. Último acesso: 30/01/2016.
- Rafael Fernandes Novo, Jose Manoel Souza das Neves, M. M. d. A. **O crescimento do big data e as possíveis implicações éticas do seu uso na análise das redes sociais**, out 2014.
- Olavsrud, T. **Big data predictions for 2014.** <http://www.cio.com/article/2369764/big-data/132163-12-Big-Data-Predictions-for-2014.html>, 2013. Último acesso: 25/01/2016.
- Power, D. J. **A brief history of decision support systems.** <http://dssresources.com/history/dsshistory.html>, 1944. Último acesso: 24/01/2016.

- Press, G. **A very short history of big data.** <http://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/#7e94ea9855da>, mai 2013. Último acesso: 23/01/2016.
- Elvia Nunes Ribeiro, Gilda Aquino de Araujo Mendonça, A. F. d. M. **A importância dos ambientes virtuais de aprendizagem na busca de novos domínios da ead,** abr 2007.
- Ribeiro, C. J. S. **Big data: os novos desafios para o profissional da informação,** 2014.
- Janessa Rivera, R. v. d. M. **Impact of iot on business at the gartner symposium.** <http://www.cio.com/article/2369764/big-data/132163-12-Big-Data-Predictions-for-2014.html>, 2014. Último acesso: 25/01/2016.
- Rocha, C. **é preciso uma ética para o big data.** <http://blogs.estadao.com.br/link/e-preciso-uma-etica-para-o-big-data/>, jun 2015. Último acesso: 16/02/2016.
- de Araujo Rosa, R. **Big data: conceitos e os 5vs.** <https://www.linkedin.com/pulse/big-data-conceitos-e-os-5vs-rodrigo-de-araujo-rosa?trk=prof-post>, jul 2015. Último acesso: 30/01/2016.
- Sasaki, C. **Tecnologia na educação.** <http://revistaescola.abril.com.br/blogs/tecnologia-educacao/2015/11/24/voce-sabe-o-que-e-big-data/>, nov 2015. Último acesso: 14/02/2016.
- Pasqueline Dantas Scaico, Ruy Jose G. B. de Queiroz, A. S. **O conceito big data na educação,** 2014.
- Schmarzo, B. **What universities can learn from big data â higher education analytics.** https://infocus.emc.com/william_schmarzo/what-universities-can-learn-from-big-data-higher-education-analytics/, jul 2014. Último acesso: 14/02/2016.
- Setzer, V. W. **Dado, informação, conhecimento e competência.** <http://www.ime.usp.br/~vwsetzer/dado-info.html>, mai 2015. Último acesso: 24/01/2016.
- Taurion, C. **Big data = volume + variedade + velocidade de dados.** https://www.ibm.com/developerworks/community/blogs/ctaurion/entry/big_data_volume_variedade_velocidade_de_dados?lang=en, 2012. Último acesso: 30/01/2016.
- Taurion, C. **Cientista de dados: o profissional de big data.** 9. https://www.ibm.com/developerworks/community/blogs/ctaurion/entry/cientista_de_dados_o_profissional_de_big_data?lang=en, out 2013. Último acesso: 16/02/2016.
- Taurion, C. **Lições sobre big data.** <http://cio.com.br/tecnologia/2015/03/23/licoes-sobre-big-data>, mar 2015. Último acesso: 30/01/2016.
- dos Santos, F. U. **O que é big data e como funciona?** <http://www.professionaisti.com.br/2015/03/o-que-e-big-data-e-como-funciona/>, mar 2015. Último acesso: 14/02/2016.