

Inovação na inteligência analítica por meio do *Big Data*: características de diferenciação da abordagem tradicional

RAFAEL NOVO

Centro Estadual de Educação Tecnológica Paula Souza - SP - Brasil
rfnovo@hotmail.com

JOSÉ MANOEL SOUZA DAS NEVES

Centro Estadual de Educação Tecnológica Paula Souza - SP - Brasil
jmneves.fatec@gmail.com

Resumo - Os sistemas de inteligência analítica lançam mão de um conjunto de tecnologias e processos com o objetivo de gerar diferencial competitivo para as empresas. Entretanto, a multiplicação do volume de informação existente, aliado ao surgimento de novas fontes de dados não estruturados, representam um desafio ao modelo tradicional de análise de dados. Nesta esteira, o termo *big data* foi cunhado para definir um conjunto de tecnologias e métodos diferenciados para lidar com a análise de dados, permitindo as empresas evoluírem seu modelo analítico para continuar a competir de forma eficiente. Este artigo provê uma breve contextualização do *big data*, seu valor para as empresas e como estas estão utilizando esta tecnologia inovadora.

Palavras-chave: *Big data*; competitividade; inteligência analítica; inovação.

Abstract - *Analytical Intelligence systems uses a combination of technologies and processes with the objective of generating competitive advantage for business. However, the amount of volume is multiplying and together with the appearance of new unstructured data sources, it represents a challenge to the traditional data analysis. The term big data defines a set of new technologies and methods for data analysis, enabling companies evolve their analytical model to continue to compete effectively. This article provides a brief background of big data, their value to enterprises and how they are using this innovative technology.*

Keywords: *Big data*; competitiveness; analytics; innovation.

1. Introdução

A inteligência analítica está presente em empresas dos mais diversos segmentos, seja no cerne do negócio de multinacionais como Google ou Walmart, seja oferecendo *insights* para áreas específicas como marketing e finanças de empresas de diversos segmentos. A inteligência analítica permite melhorar o

poder de competição das empresas, além de ser uma parte importante dos sistemas de BI (*business intelligence*¹)(DAVENPORT, 2007).

Objetiva-se aqui oferecer uma breve conceituação do *big data*, termo cunhado para definir um modelo evolutivo de inteligência analítica. É também definido em pesquisa realizada por Casonato (2013, p. 4) como “ativos de informações com alto volume, velocidade e variedade que demandam formas eficientes e inovadoras de processamento dos dados para obtenção de insights melhorados e tomadas mais precisas de decisão”².

Brynjolfsson (2012) afirma que soluções de *big data* possuem um potencial ainda maior de beneficiar as empresas do que as soluções analíticas tradicionais, as quais já são utilizadas e conhecidas por abrir novas oportunidades e oferecer vantagens competitivas para aquelas que as utilizam. O autor, por considerar que *big data* trará uma revolução na gestão, afirma que os executivos precisam encarar este novo modelo o quanto antes.

É descrito nessa pesquisa um conjunto de estudos técnicos e acadêmicos apontando a crescente importância do *big data* como uma das tecnologias estratégicas para as empresas, pontuando com exemplos de instituições que obtiveram vantagens competitivas.

2. Referencial Teórico

Big data foi listado no estudo dirigido por Cearley (2013) como uma das 10 tendências de tecnologias estratégicas com o potencial de afetar a vida das pessoas e das empresas. Esta lista foi montada com base no potencial de cada tecnologia por serem altamente disruptivas ou por estarem na eminência de sofrer uma aceleração, mudança ou ponto de inflexão, tornando a tecnologia estratégica e aplicável a um amplo mercado nos próximos 2 anos.

Cearley (2013) avisa que, embora este rol tenha sido elaborado considerando o potencial de impacto dessas tecnologias em uma ampla gama de empresas, cada companhia deve individualmente criar sua própria lista, considerando fatores particulares como necessidades de negócio e modelo de adoção de novas tecnologias.

Bughin (2011) relata que muitos executivos só procuram por tecnologias de *big data* quando tem uma necessidade urgente de responder a uma ameaça ou identificam uma oportunidade de atacar um determinado nicho de mercado. No entanto, identificar essas oportunidades e ameaças pode ser difícil. O autor cita uma empresa de telecomunicações europeia que após aumentar sua fatia de participação no mercado decidiu aproveitar o momento criando um sistema que permitisse identificar os motivos que poderiam levar seus clientes a optar por uma ou outra empresa. Para isso foi necessário analisar uma quantidade grande e em constante crescimento de informações disponibilizadas pelos próprios usuários em mídias sociais e canais de comunicação web da empresa.

Porém a implementação intempestiva de soluções de *big data* é um erro, como descrito por Bughin (2011). É necessário que a sua estratégia de uso esteja intimamente conectada a estratégia da empresa e para isso é necessário um extenso planejamento dos executivos para decidir como os recursos analíticos

¹ Sistemas de inteligência de negócio são comumente chamados pelo nome em inglês de *business intelligence* ou simplesmente BI.

² Tradução livre.

devem ser aplicados para obter o resultado esperado. Em alguns casos, a melhor decisão pode ser esperar antes de embarcar em um projeto como este. A empresa pode estar se reestruturando, redefinindo o seu foco estratégico ou passando por dificuldades financeiras. Entretanto, é importante que os executivos tenham uma visão estratégica de longo prazo e acompanhem de perto a concorrência, a qual pode desenvolver avanços na competição analítica, obtendo vantagem competitiva e colocando em risco o plano traçado.

O *big data* é capaz de trazer *insights* importantíssimos para a empresa, mas a decisão e visão do executivo sempre será essencial para o negócio, porém agora, muito mais embasada em dados. Como no exemplo trazido por Vivek Ranadive, CEO da Tibco³ e um conhecido autor de livros sobre o tema, em uma determinada loja varejista de material de jardinagem, sabe-se que compradores de semente tem 90% de probabilidade de comprar fertilizante e 40% de chance de comprar móveis de jardim. Em uma primeira análise pode-se imaginar que a melhor estratégia esteja em incentivar a compra de fertilizante ou colocar este perto das sementes. Porém analisando mais a fundo, pode-se determinar que a melhor estratégia é desenvolver ações que incentivem a venda dos móveis! Afinal, há uma margem maior de crescimento para este produto (GALLANT, 2011).

É comum a comparação – e algumas vezes a falta de uma definição mais clara e objetiva – entre *big data* e a inteligência analítica tradicional, pois ambos tem a capacidade de analisar dados e convertê-los em informação para gerar vantagem competitiva. Há, contudo, três grandes diferenças listadas por Brynjolfsson (2012) e corroboradas por Casonato (2013): volume; velocidade e variedade. Essas características, como exemplificado adiante, não aparecem de forma isolada. São constantes os exemplos de soluções de *big data* abordando problemas de volume e velocidade de análise, ou velocidade e variedade ou os três fatores em conjunto.

A quantidade de dados existente tem crescido de forma acelerada. Segundo Brynjolfsson (2012), em 2012 foram criados diariamente cerca de 2,5 exabytes⁴ e o volume de dados trafegado a cada segundo na internet neste mesmo ano foi maior do que toda a informação que estava armazenada na rede há 20 anos atrás. Isso cria uma nova oportunidade para as empresas analisarem um volume muito maior de dados que podem conter informações relevantes para melhorar seus processos de análise de dados, gerando uma percepção muito mais apurada de seu conhecimento do mercado, modelo de relacionamento com clientes e outros aspectos corporativos que permitam a estas empresas melhorar sua competitividade.

Esta necessidade de analisar grandes volumes de dados aumentará nos próximos anos, visto que a quantidade de informações existente, chamada de universo digital, segundo estudo de Gantz (2012), crescerá em até 50 vezes entre os anos de 2010 e 2020 e totalizará um crescimento de até 300 vezes se considerado os anos de 2005 a 2020, saindo de 130 exabytes para aproximadamente 40 zetabytes⁵.

Esse crescimento deve-se a diversos fatores, explicados por Brynjolfsson (2012, p. 2), como o surgimento de novas fontes de informação oriundas de equipamentos capazes de gerar grandes quantidades de dados que podem ter

³ Empresa americana de capital aberto, que fornece soluções de análise de dados

⁴ 1 exabyte = 10⁶ terabytes ou 10¹⁸ bytes (considerando base decimal).

⁵ 1 zetabyte = 10³ exabytes (considerando base decimal)

relevância para as empresas, tais como “telefones celulares, compras na internet, redes sociais, comunicação eletrônica, GPS e maquinário computadorizado” que “produzem torrentes de dados como subproduto de sua operação normal”.

O segundo aspecto de diferenciação do *big data* é a velocidade de processamento das informações, que permite a análise imediata, ou quase, dos dados, possibilitando à empresa ser muito mais ágil do que seus concorrentes, trazendo perspectivas rápidas que podem, por exemplo, dar uma óbvia vantagem competitiva para analistas do mercado financeiro e gestores de empresas, ou ainda permitir que uma rede de supermercados crie ofertas de marketing em tempo real para seus consumidores (BRYNJOLFSSON, 2012).

Brown (2011) afirma que as lojas líderes do varejo estão constantemente executando experimentos para identificar fatores que levem a maior participação dos clientes e conseqüentemente maiores vendas. Embora muitas empresas já façam estes experimentos há algum tempo, o *big data* pode levá-los a um próximo nível. A futura geração de lojas de varejo será capaz de identificar comportamentos individualizados de cada cliente, utilizando, por exemplo, o *clickstream*, que é a sequência de cliques de mouse em sua página, determinando o padrão de navegação, se adaptando as preferências do cliente em tempo real. Neste modelo é possível identificar quando um cliente está próximo de efetuar uma compra, permitindo a empresa influenciar na decisão do cliente, por meio de uma promoção relâmpago, oferecendo um pacote com outro produto que cliente esteja interessado, ganhos extras em programas de milhagem ou outra oferta que tenha sido identificado como relevante para o cliente e que seja vantajoso à empresa. Desta forma é possível não apenas garantir a venda como agregar produtos com maiores margens para a empresa.

O terceiro item listado, a variedade, refere-se às diversas fontes de dados, tais como: mensagens; leituras de sensores; câmeras de segurança; sinais de GPS de celulares e outros, que podem ser usados para análise e posterior extração de informação pertinente ao negócio. Fontes para o *big data* podem englobar tudo, desde dados de voz de *call center* até dados de genôma de pesquisa biológica e medicinal (DAVENPORT, 2012) e pode ser provenientes das mais diversas origens, como uma copiadora ou até mesmo um motor de jato (BROWN, 2011). Nota-se que algumas das fontes de dados de maior importância que podem ser utilizadas por sistemas de *big data* são relativamente novas, como as redes sociais, dentre as quais se destacam o Facebook lançado em 2004 e o Twitter em 2006, que aliadas a também recente proliferações de smartphones, tablets, ultrabooks, redes wireless e redes de dados 3G e 4G, permitem às pessoas estarem conectadas em praticamente qualquer lugar do globo, gerando conteúdo, ideias e opiniões.

O poder destas mídias sociais merece um destaque especial, pois são uma poderosa fonte de dados à ser utilizada pelas empresas nos seus projetos de *big data*. Brynjolfsson (2012) cita o uso de mensagens do Twitter que foram analisadas e conseguiram antever em duas semanas os boletins oficiais de propagação da cólera no Haiti após o terremoto em 2010. O autor menciona ainda o estudo realizado por Erik e Lynn Wu que usaram dados obtidos livremente em buscas na internet para prever a variação média do mercado imobiliário dos EUA (Estados Unidos da América), conseguindo obter estimativas mais precisas do que a oficial, da Associação Americana dos Corretores de Imóveis⁶, que utilizava

⁶ Tradução Livre

um modelo complexo, porém alimentado por dados históricos que não conseguiam acompanhar a nova dinâmica do mercado.

Como relatado por Hea (2013), as mídias sociais já são utilizadas por diversas empresas, em especial para propaganda, campanhas de marketing e vendas, monitoração de satisfação dos clientes e até mesmo para desenvolvimento ou inovação de produtos. Exemplifica-se o exposto com a cadeia de hotéis Starwood⁷ que há alguns anos usam as redes sociais para estarem mais próximos de seus hóspedes, resolvendo rapidamente qualquer percalço que possa ter ocorrido nas suas estadas nos hotéis ou até mesmo para fazer sugestões de novas viagens.

O *big data* em conjunto com as redes sociais permite o que o Brown (2011) chamou de técnicas de análise de sentimento, isto é, fazendo uso da análise de quantidades imensas de dados criados pelos clientes em mídias sociais, blogs e outras fontes na internet, obtêm um feedback real do cliente com relação ao atendimento prestado, qualidade do produto ou impacto de uma campanha de marketing. A análise e obtenção deste feedback imediato permite que a empresa resolva o problema ou que a área de marketing ajuste sua campanha rapidamente, maximizando o resultado.

Todas estas fontes de dados podem ser transformadas em informação valiosas, tais como gostos, hábitos, modo de consumo e satisfação das pessoas, a serem utilizadas pelas empresas como diferencial competitivo, como descrito por Davenport (2012).

A grande maioria destes novos dados – que estão tornando-se diferencial competitivo na análise de dados – são do tipo não estruturado, ou seja, não estão organizados em tabelas e colunas de um banco de dados tradicional. Por exemplo, o Google processa diariamente aproximadamente 24.000 terabytes de dados e apenas uma ínfima porção destes dados está armazenado em bases de dados tradicionais (DAVENPORT, 2012). Mesmo nos dados provenientes de dentro das empresas, Hea (2013) coloca que 80% destes também não são estruturados, como por exemplo, e-mails, documentos, entre outros.

Além de não estruturadas, estas novas fontes de dados possuem outra característica marcante: a volatilidade. São informações que, conforme descritas por Davenport (2012), não estão guardadas em um banco, elas estão em movimento constante, sendo alteradas a cada momento, tais como o fluxo de cliques dos consumidores em um site, ou os relacionamentos de potenciais clientes em sites como Facebook ou LinkedIn.

Dentre as diversas tecnologias que compõe o *big data*, o chamado *text mining* ou mineração de texto é responsável pela análise destes dados não estruturados. A análise de textos livre possui uma complexidade superior à análise de dados estruturados e o seu principal objetivo é encontrar padrões para conseguir extrair informações e inteligência analítica destes textos (HEA 2013). Este é um tema amplo que não será aprofundado neste artigo, na medida que extrapola o âmbito da presente pesquisa.

Torna-se necessário que as empresas passem a pensar e considerar o fluxo contínuo de informação e processos, ao invés de olhar apenas para dados que ocorreram no passado. Tom Deutsch, diretor do programa de tecnologias de *big data* e analítica aplicada na IBM ressalta: “Análise durante a transmissão

⁷Cadeia norte-americana com milhares de hotéis espalhados pelo mundo. No Brasil sua marca mais conhecida são os hotéis da Sheraton.

permite o processo dos dados durante um evento, melhorando o resultado”⁸ (DAVENPORT, 2012, p. 2).

O *big data* abre a possibilidade de um modelo de decisão diferente, pois permite às empresas criação de experimentos controlados para testar hipóteses que guiarão a tomada de decisão em, por exemplo, novos investimentos ou mudanças operacionais. Possibilitando centenas ou milhares de experimentações é possível distinguir entre simples correlação de eventos daqueles que realmente possuem uma ligação de causa e efeito (BROWN, 2011).

Vivek ressaltou em sua entrevista para Gallant (2011) a importância da análise de eventos e da capacidade de criar sistemas preditivos. Conseguir inferir o que pode acontecer com um mínimo de antecedência é mais importante que pode analisar toda a informação global, porém 6 meses depois do ocorrido. Prever a intenção de um cliente de mudar-se para a concorrência tem muito mais valor do que analisar porque o cliente deixou a empresa. Não que entender os motivos que levaram a esta decisão do cliente não tenha seu valor, mas poder tomar ações de contenção, impedindo a saída do cliente, é mais importante do que agir reativamente, melhorando o que estava errado ou investindo em campanhas caras para tentar trazer o cliente perdido de volta. Esta mesma visão é compartilhada por Hea (2013) que adiciona a importância do uso de dados disponíveis em mídias sociais para a capacidade de realizar estas predições.

Apesar do impacto positivo do *big data*, Cearley (2013) adverte que novas tecnologias possuem diversos vetores de impacto, podendo causar a disrupção de outras tecnologias tradicionais como *data warehousing*⁹ e gerenciamento da informação. Contudo, essa disrupção não precisa, necessariamente, ser algo ruim, pois se explorada corretamente, pode trazer benefícios para a empresa.

Um dos desafios enfrentados pelas empresas é o compartilhamento de informação entre unidades de negócio, criando algumas vezes verdadeiros feudos de dados. Brown (2011) cita as empresas financeiras como exemplo deste modelo em que a falha de compartilhamento de informação entre mercados financeiros, gerenciamento das finanças, empréstimos e outros impede a criação de uma visão completa de seus clientes e o entendimento das possíveis ligações entre os diversos segmentos.

Vivek, em entrevista para Gallant (2011) pontua muito bem a evolução do modelo de análise de dados tradicional do século XX para um novo modelo do século XXI, comparando-o com o próprio comportamento humano. O ser humano consegue processar uma quantidade grande de informações e tomar decisões em tempo real. No exemplo usado por Vivek, quando uma pessoa está dirigindo ela precisa estar atenta à diversas fontes de informação, processando todas em tempo real para tomar a melhor decisão. Se identificar um obstáculo na via, precisa responder imediatamente, pois de nada adianta querer desviar depois de já ter atingido o obstáculo. Essa metáfora pode ser aplicada para o uso corporativo de soluções de *big data*, que consegue analisar em tempo real quantidades massivas de informações, advindas de diversas fontes.

⁸ Tradução Livre

⁹Repositório central de dados, criado pelas empresas para armazenar dados oriundos de diversas bases de dados espalhadas pela organização, com o objetivo de permitir a análise destes dados de forma mais fácil e eficiente.

3. Metodologia

A pesquisa descreve o uso do *big data* pelo mercado, lançando mão de fontes de dados secundárias com o objetivo de aprofundar o conhecimento sobre o tema. A pesquisa bibliográfica ou revisão de literatura é para Roesch (2007) uma das partes mais longas e trabalhosas do projeto, pois requer seleção, leitura e análise para determinar os textos relevantes ao tema abordado.

Buscou-se utilizar principalmente estudos e pesquisa de institutos renomados como Gartner e McKinsey, além de artigos científicos contendo fundamentação teórica sobre o tema e estudos de casos de empresas ou segmentos empresariais que estão utilizando o *big data*. Realizando a agregação destas fontes de dados foi possível discorrer sobre o tema com um embasamento de casos práticos.

4. Resultados e discussões

A transformação de dados crus em informação relevante ainda é um processo extremamente complexo, especialmente para grandes volumes de dados e por este motivo muitas empresas restringem o escopo desta transformação em uma parcela muito pequena da massa de dados disponível. Como visto no estudo de Gantz (2012), existe uma lacuna muito grande entre a quantidade de dados considerada útil para as empresas e o volume atual de dados efetivamente classificados ou marcados. Existem duas lacunas aqui que podem ser preenchidas ao longo dos próximos anos com adoção de sistemas de *big data* pelas empresas. A primeira é identificar as informações relevantes, dentro de toda a massa de dados disponível. Já a segunda consiste em classificar e analisar estes dados, visando obter mais insights e uma melhor assertividade.

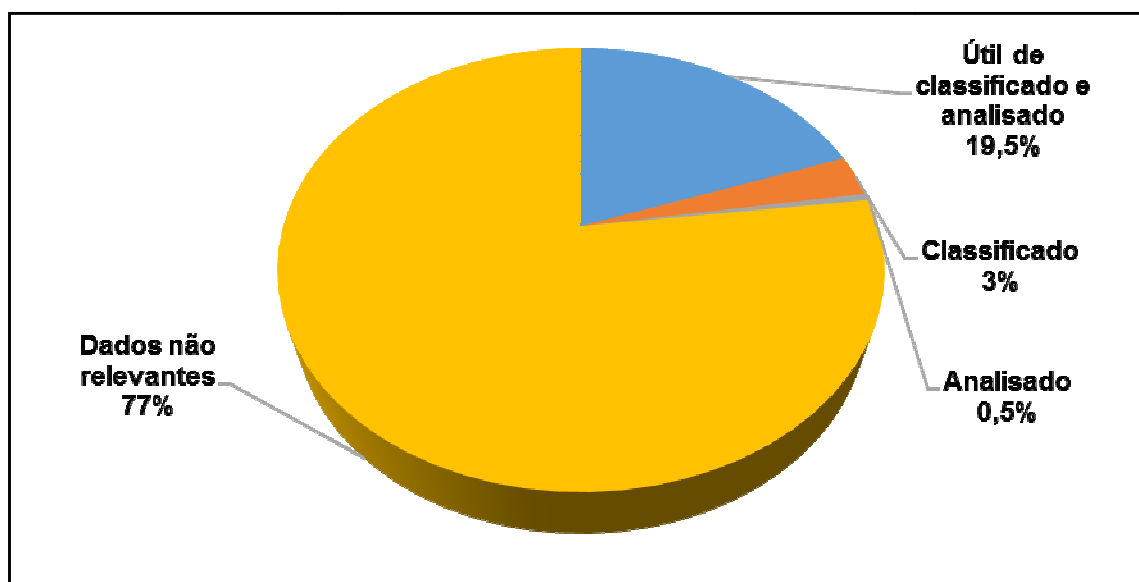


Figura 1: A lacuna de dados úteis, porém não classificados e analisados

Fonte: Adaptado de Gantz(2012, p. 11)

Porém existe um potencial para obtenção de benefício na exploração desta massa de dados ainda intocada, como levantado por Brown (2011). A Figura 2 compara o potencial valor para cada segmento com a dificuldade em viabilizar projetos de *big data*. Para determinar o potencial de benefício foi considerado fatores como: turbulência de mercado, estrutura, quantidade de transações, número de clientes potenciais, parceiros de negócios e quantidade de dados disponíveis. Enquanto que para calcular a facilidade de captura dos dados foi avaliado o número de empregados com conhecimento analítico no segmento de indústria, investimento médio em TI, facilidade de acesso às fontes de dados e o grau de uso de dados nas decisões gerenciais.

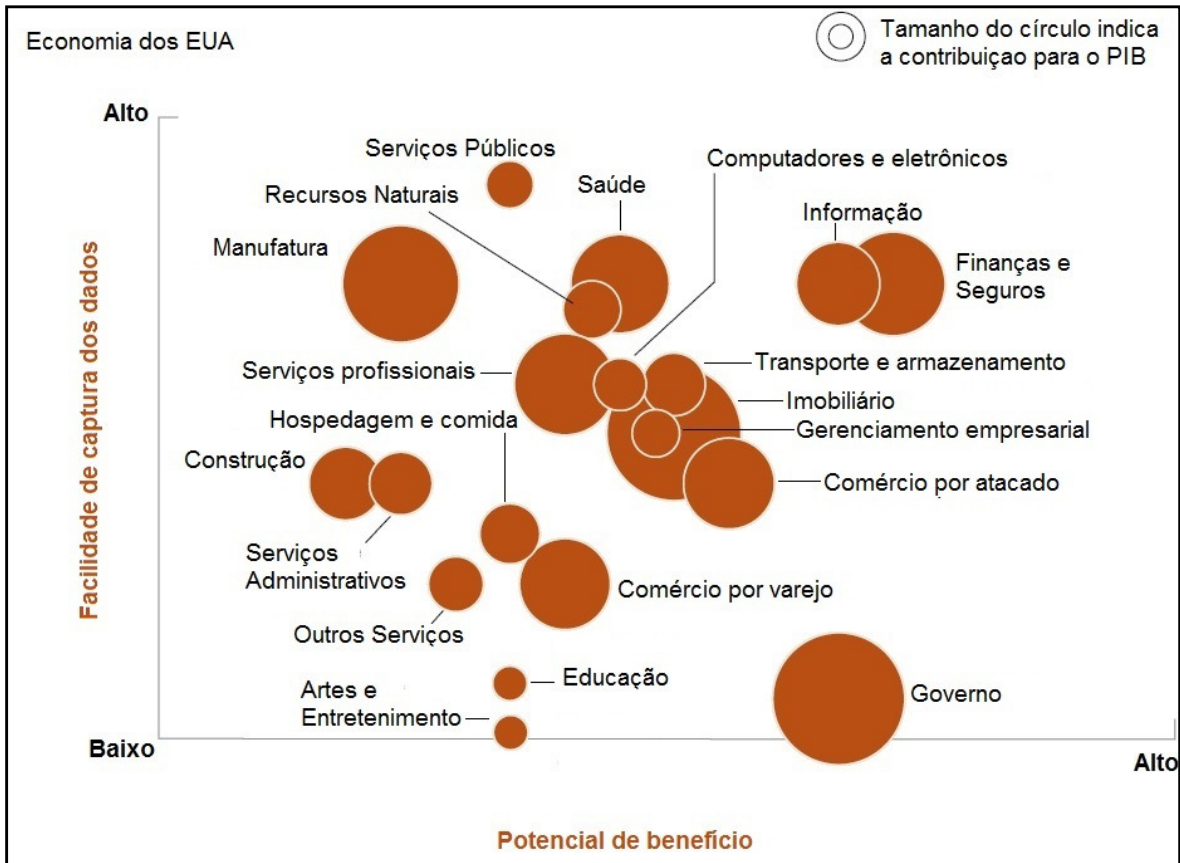


Figura 2: Facilidade de uso versus valor ao negócio
Fonte: Adaptado de Brown (2011, p. 6)

Existe uma série de exemplos de empresas dos mais diversos segmentos, utilizando o *big data* para melhorar seus processos produtivos, desenvolver novos produtos ou conhecer melhor o seu mercado consumidor.

Brynjolfsson (2012) descreve um estudo feito no MIT Center for Digital Business objetivando comprovar a hipótese de melhor desempenho das empresas movidas a dados. Para isto foram aplicadas entrevistas estruturadas com executivos de 330 empresas norte-americanas de capital aberto sobre suas práticas de gestão organizacional e TI (Tecnologia da Informação), agregando o resultado destes a dados de desempenho disponíveis em balanços anuais e fontes independentes. Deste estudo foi identificado que as empresas que mais faziam uso de dados nas tomadas de decisão, aproximadamente um terço das

empresas entrevistadas, possuíam, em média, 5% mais produtividade e uma rentabilidade 6% maior do que as demais.

Comparando os exemplos aqui listados é possível identificar as três dimensões do *big data* (volume, velocidade e variedade) presentes. Nota-se que em alguns cenários descritos é aplicado apenas uma das dimensões enquanto outros utilizam as duas ou três dimensões simultaneamente.

Bughin (2011) traz o exemplo de uma empresa europeia de telecomunicações, na qual foi criado um comitê com membros das diversas áreas da empresa. Este comitê supervisionou o projeto de *big data* para direcionar os experimentos à responder perguntas como: “Quão competitiva as nossas marcas aparecem nas mentes dos nossos clientes no momento que eles fazem uma decisão de compra?”¹⁰ e “Quais os fatores que mais importam para os usuários e quão bem posicionado nós estamos para comunicar estes fatores aos nossos clientes?”¹¹. Com base nestas perguntas foram analisadas diversas fontes de dados para obter algumas revelações interessantes, como por exemplo, os clientes seriam melhor influenciados a fechar um pacote de serviços (TV, Internet e telefonia) de acordo com a programação do serviço de TV e menos influenciados por características dos serviços de voz. Essa foi uma informação contraditória da obtida por meio de questionários tradicionais e permitiu que a empresa mudasse sua campanha de marketing. Esse exemplo ressalta que para os dados serem úteis eles precisam atravessar as barreiras organizacionais, o que é, geralmente, difícil e causa brigas e disputas internas por poder corporativo.

A Sears¹² era uma das empresas que sofria com o crescimento da massa de dados. Necessitando desenvolver promoções personalizadas para seus clientes, o processo de analisar o gigantesco volume de informações que possuía levava cerca de 8 semanas, tornando as promoções, muitas vezes, obsoletas e ineficientes. Utilizando tecnologias de *big data* e promovendo a unificação das bases de dados, o tempo exigido para análise caiu para apenas 1 semana (BRYNJOLFSSON 2012).

Brown (2011) traz 5 exemplos de empresas que usaram dimensões diferentes do *big data* para obter vantagem competitiva de diversas formas, como melhorando seus processos, produtivos, diminuindo estoques e conseqüente custos operacionais, aumentando suas vendas através da melhora do conhecimento dos seus clientes ou criando novos produtos.

Uma grande empresa de petróleo que conseguiu reduzir os custos com pessoal em quase 25% e ao mesmo tempo aumentar a produtividade em 5%. Este ganho de eficiência operacional foi possível utilizando sensores ao longo da linha de produção, enviando dados que eram analisados em tempo real permitindo ajustes imediatos na linha, evitando assim intervenções manuais, que são, em geral, demoradas, caras e perigosas.

Uma empresa de bebidas europeia reuniu informações sobre a previsão do tempo ao seus dados de estimativa de demanda para diminuir o inventário e melhorar a assertividade nas estimativas de consumo em aproximadamente 5%, analisando dados como previsão de chuva, temperatura e número de horas de sol em um determinado dia.

O McDonalds que equipou alguma lojas com dispositivos que rastreiam dados operacionais, interações com os clientes, tráfego nas lojas e padrão de

¹⁰ Tradução livre

¹¹ Tradução livre

¹² Gigante do setor do comércio de varejo norte-americana, com uma rede de lojas de departamentos espalhadas por todo o país.

ordens, permitindo a análise dos impactos na produtividade e nas vendas, gerados por mudanças em itens como cardápio, atendimento e desenho da loja.

Cita-se também o caso de uma empresa líder do varejo norte-americano, a qual conduziu experimentos utilizando dados de movimentação das lojas em conjunto com registros de vendas para criar um modelo que permitisse encontrar o melhor posicionamento dos produtos na loja e o preço ideal para estes produtos. Como resultado esta empresa conseguiu reduzir o estoque em 17% e ao mesmo tempo maximizar a venda de produtos de sua própria marca, aumentando os lucros sem comprometer a sua porcentagem de participação no mercado.

Uma empresa global desenvolveu um novo produto, após utilizar o *big data* para melhorar o seu processo de manufatura, eles aprenderam tanto sobre os próprios processos que decidiram prestar consultorias para outras empresas em situações semelhantes. Atualmente tal estabelecimento agrega dados de chão de fábrica e cadeia de suprimentos de diversas empresas e vende ferramentas de software para melhorar a performance destas empresas. Este novo produto/serviço baseado no *big data* atualmente representa uma renda maior do que sua manufatura.

Davenport (2012), por sua vez, cita empresas de cartão de crédito que desejavam encontrar os melhores prospectos de clientes em grandes bases de dados, porém demoravam semanas para extrair, preparar e elaborar ofertas. Atualmente, com a monitoração das interações no *site web* e no *call center*, as empresas conseguem criar ofertas personalizadas em menos de um segundo, além de otimizá-las constantemente por meio da mensuração dos resultados.

Um dos exemplos mais conhecidos de empresa que faz uso intensivo da análise de dados é o Walmart, comentado por Davenport (2007) e detalhado por Hays (2004). Este gigante do varejo utiliza técnicas de análise de dados para melhorar sua competitividade de várias formas:

- Entendendo melhor o padrão de consumo dos seus clientes, correlaciona o acontecimento de eventos à compra de produtos alimentícios, exemplificado Linda M. Dillman, CIO do Walmart, em entrevista a Hays (2004): “No passado nós não sabíamos que as vendas de *Pop-Tarts*¹³ de morango aumentavam em até 7 vezes antes de um furacão”¹⁴ e complementa “o item mais vendido antes de um furacão é cerveja”¹⁵. Tal *insight* permite ao Walmart garantir o abastecimento correto das lojas, a satisfação dos clientes e um maior lucro.
- Munir os seus fornecedores de informação, melhorando o gerenciamento de seu *supplychain*¹⁶, garantindo um melhor abastecimento das lojas, reduzindo os seus custos e melhorando sua competitividade.

Quando Hays (2004) escreveu seu artigo para o jornal New York Times os volumes estimados de dados armazenados pelo Walmart referente aos seus clientes e padrões de consumo era de aproximadamente 460 terabytes de dados, armazenados em um sistema da empresa Teradata, que hoje se autoproclama líder em soluções de *big data*. Davenport (2007) apresentou uma estimativa de 583 terabytes armazenados pelo Walmart e recentemente Brynjolfsson (2012) trouxe a estimativa de 2,5 exabytes de dados armazenados dentro dos sistemas

¹³ Pop-Tart é uma espécie de bolo industrializado.

¹⁴ Tradução Livre

¹⁵ Tradução Livre

¹⁶ Cadeia de suprimentos é comumente referenciada por seu equivalente em inglês.

de *big data* do Walmart, com informações diversas sobre os seus clientes, histórico financeiro e padrões de compras.

Uma empresa de telecomunicações da Índia estava adicionando à sua base uma média de 5 milhões de novos assinantes por mês, entretanto estavam perdendo 1,5 milhões de clientes para a concorrência todos os meses. Após uma análise minuciosa, identificaram que os clientes tinham uma alta tendência de mudar de operadora depois de 6 chamadas interrompidas. Desta forma a empresa, após a quinta chamada interrompida, disparava um torpedo com uma promoção para o cliente, evitando assim a sua fuga para a concorrência. Apenas com um modelo de *big data* foi possível a análise de bilhões ou até mesmo trilhões de registros ou eventos, que permitiram a empresa criar um modelo individualizado de relacionamento com seus clientes (GALLANT, 2011).

Mais uma evidência do aumento da competitividade que uso do *big data* pode trazer para os mais diversos segmentos é o estudo de caso realizado por Bughin (2011) com a AstraZeneca, uma gigante do ramo farmacêutico. Mark Lelinski, vice-presidente da empresa relata que antigamente as principais preocupações da área de pesquisa e desenvolvimento era garantir que um novo medicamento era eficiente, seguro e aprovado pelas regulamentações. Recentemente uma nova preocupação surgiu, tão importante quanto as demais: como esse produto pode ser comparado com os concorrentes e por qual motivo ele deve ser escolhido frente a concorrência cada vez mais acirrada.

Este problema levou a criação de um modelo de *big data* que concatenaram diversas fontes de dados para medir o custo total para tratar uma determinada doença, incluindo idas ao médico, exames diagnóstico, atendimentos de emergência, entre outros. Os medicamentos representavam apenas 11% deste custo total, por isso, se uma droga conseguisse reduzir a fatia de remanescente de custos de 89%, esse poderia ser um diferencial e uma forte justificativa financeira para sua aquisição. Importante lembrar que os clientes da AstraZeneca não são somente os consumidores finais, mais também planos de saúde e hospitais. Mark destaca que o apoio da alta administração foi fundamental para viabilizar esse projeto.

Com este novo modelo em funcionamento é possível determinar, ainda no começo do ciclo de desenvolvimento, a viabilidade comercial desta nova droga, possibilitando interromper o que seria um grande investimento com baixo retorno, e focar o investimento em pesquisas e desenvolvimentos de drogas que sejam lucrativas para a empresa e que melhorem a vida dos pacientes reduzindo o custo total de atendimento.

5. Considerações finais

Como visto, as empresas estão nadando em um mar de dados, os quais são muito volumosos ou muito complexos para serem analisados da forma tradicional (DAVENPORT 2012). E a análise dessa gigantesca massa de dados está rapidamente tornando-se a nova fronteira da competição empresarial (BUGHIN, 2011).

Por conseguinte, *obig data* surge 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. Empresas que já utilizam a análise de dados como ferramenta de competitividade largam com vantagem

nessa corrida analítica, a se destacar a melhoria nos processos produtivos, de gestão, de operação ou nos seus modelos de vendas e interação com os consumidores.

Como em todo modelo competitivo, ganha mais vantagem quem chega primeiro, daí a importância das empresas encararem os dados como um ativo primordial e criarem projetos para investigação e subsequente implementação do *big data* como ferramenta de diferenciação no mercado, sempre alinhados aos objetivos corporativos e com metas bem estabelecidas.

Este estudo inicial sobre o tema abre via para diversas pesquisas: aprofundamento nos componentes do *big data*; estudo bibliométrico para medição de produção científica relacionada; melhoria das dimensões da TI para endereçar as características do *big data* (velocidade, volume, variedade); estudos de caso de empresas; implicações éticas na obtenção e uso de informações detalhadas dos consumidores por meio desta nova tecnologia.

Referências

BROWN, B; CHUI, M; MANYIKA, J. Are you ready for the era of 'big data'?. *McKinsey Global Institute*. 2011

BRYNJOLFSSON, E; MCAFEE, A. Big Data - A Revolução da Gestão. *Harvard Business Review*. 2012

BUGHIN, J; LIVINGSTON, J; Marwaha, S. Seizing the potential of 'big data'. *McKinsey Global Institute*. 2011

CASONATO, R; *et al.* Top 10 Technology Trends Impacting Information Infrastructure, 2013. *Gartner*. 2013

CEARLEY, D; Claunch, C. The Top 10 Strategic Technology Trends for 2013. *Gartner*. 2013

DAVENPORT, T; BARTH, P; BEAN, R. How 'Big Data' is Different. *MIT Sloan Management Review*. 2012

DAVENPORT, T; HARRIS, J. *Competição Analítica: Vencendo Através da Nova Ciência*. Campus, 2007.

GALLANT, J. TIBCO CEO: How Real-Time Computing Will Change the Landscape. *ComputerWorld*. 2011

GANTZ, J; REINSEL, D. THE DIGITAL UNIVERSE IN 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. *IDC*. 2012

HAYS, C. What Wal-Mart Knows About Customers' Habits. *The New York Times*. 2004

HEA, H; ZHAB, S; LI, L. Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management*. 2013

ROESCH, S. *Projeto de estágio e de pesquisa em administração*. 3ª ed. São Paulo: Atlas, 2007