

Big Data, ameaça ou privilégio para a sociedade?

Mariana Tessmann Martins * Nicolas Rodio †

Universidade Federal de Santa Catarina - UFSC. R. Eng. Agrônomo Andrei
Cristian Ferreira, s/n - Trindade, Florianópolis - SC, 88040-900

Resumo

Durante todo o decorrer da história, o armazenamento de dados possibilitou o desenvolvimento da sociedade. Desde as primeiras marcas escritas até as grandes produções de dados pela internet, todos os dados possibilitam com que se tenha registro do que a humanidade já fez, que deu certo ou errado e assim ajuda nas escolhas do que vem ao futuro. O que o engenheiro tem relacionado com todos esses dados armazenados, ou o que fazer com eles de forma inteligente é um assunto que se deve entrar em pauta.

Palavras-chaves: Big Data. Ciência de dados. Armazenamento de dados. Engenharia

Introdução

Com o aumento do desenvolvimento tecnológico e a uma crescente disponibilidade de acesso à internet para sociedade nos últimos anos, nota-se que a quantidade de dados gerados e processados cresce alarmantemente. A partir de 2015, com o surgimento da internet das coisas, a sociedade passou a armazenar mais de um *zetabytes* de dados, sendo assim, devido a essa enorme quantidade de dados gerados a cada segundo, deu-se nome de *Big Data*. De acordo com (HILBERT; LÓPEZ, 2011), por volta de 2003, a humanidade começou a armazenar mais dados digitais em *CD's*, *pendrives* e *HD's*, do que em meios analógicos, como *LP's*, jornais e fotos. E atualmente os dados armazenados passam de 2,5 quintilhões de *bytes* por dia.

Devido à essa crescente produção de conteúdo, ou seja, criação de dados, e com o aumento da capacidade e poder computacional, como também uma diminuição de preço e tamanho dos *hardwares*, foi possível analisar e extrair lições dos dados, com a ciência de dados. Sendo assim, os dados guardados são utilizados para aprender novos conhecimentos, como usufruir de dados estatísticos que antigamente eram armazenados e utiliza-los para criar estratégias de *marketing*, aumentar a produtividade de uma empresa, como também reduzir custos e tomar decisões mais inteligentes. Para isso, a ciência de dados, a *Big Data* em conjunto com a engenharia aprende como manipular os dados em benefício da sociedade, e cria tecnologias como o *machine learning* e entre outras. Mas, essa exposição de informação a respeito da sociedade pode se tornar uma ameaça para ela mesma, podendo influenciar em tomadas de decisões que são importantes para a humanidade.

* <marianatessmannm8@gmail.com>

† <nicolasrodio@hotmail.com>

1 Histórico do armazenamento de dados

A saída da humanidade da pré-história começa com a escrita de tábuas de barro na antiga suméria, entretanto, inicialmente acreditava-se que a escrita acabaria com o pensamento racional do ser humano. Acreditava-se que iria se estagnar e não evoluir, pois todos iriam perder sua memória ao longo dos anos, porém o que acontece é exatamente ao contrário. Apenas com a adoção de símbolos, significados e escrita foi possível que a sociedade conseguisse ter lógicas mais complexas e desenvolvesse junto a matemática. Como é descrito em (GLEICK, 2013), imaginava-se a lógica independente da escrita, pois silogismos poderiam ser ditos tanto quanto escritos, entretanto isso não ocorre pois "a fala é passageira demais para se prestar a uma análise".

Até 1445, quando o alemão Johannes Gutenberg inventou a prensa móvel, todos os livros eram copiados a mão por escribas, a qual continham inúmeros erros. Novamente houve uma aversão a esta nova tecnologia, que poderia acabar com os empregos dos escribas além de acreditar-se que haveria tantas cópias e livros irrelevantes, que a verdadeira informação iria se perder no "mar de inutilidades". Contudo, novamente as expectativas adversas a nova tecnologia, não foram o que se concretizaram. Com a prensa de Gutenberg até o final do século XV a Europa já havia impresso mais de 20 milhões de cópias, e até o século XVIII 1 bilhão de cópias. O resultado foi idéias que começaram a circular e ser acessível a mais pessoas, autores que eram pagos e podiam investir mais no conhecimento, e principalmente o maior ganho para a sociedade, a ciência pode ser discutida, rever resultados e melhor desenvolvida.

Alan Turing certa vez exclamou: "Não, não estou interessado em desenvolver um cérebro poderoso. Busco apenas um cérebro mundano". Em (TURING, 1938) Turing acreditava que por uma linguagem formal, tudo na matemática poderia ser demonstrado. Entretanto existia números não computáveis, e o que chamamos de computadores, ainda eram as pessoas. A grande produção de dados realmente começou a ocorrer depois da invenção de Turing. Em (GLEICK, 2013) conta que Turing inventou uma máquina para computar números, fazer cálculos matemáticos. Mesmo ainda não existindo essa palavra, o que ele estava fazendo era *progamando* a sua máquina. Esta máquina poderia ver apenas um símbolo, 0 ou 1 por vez, mas também poderia armazenar esses símbolos no que ele chamava de fita. Ou seja, a Máquina de Turing é um computador digital, que lê instruções escritas como sim ou não através de 0 ou 1, e cada uma dessas instruções é chamada de bit. A partir dessas instruções ela segue os algoritmos e executa programas, a qual é o mesmo sistema que todos os computadores e sistemas binários atuais usam.

O conjunto de 8 bits que era utilizado para codificar uma letra foi denominado byte, quando foram escritos os primeiros programas. Dessa forma, os primeiros computadores trabalhavam com 8 mil bits ou 1 kilobyte. O que mudou até os dias atuais, é que os computadores adquiriram tecnologia que os possibilitou ter maior armazenamento, ou seja, maior capacidade de byte. Além disso, ficaram menores fisicamente e mais baratos.

Conforme (HILBERT; LÓPEZ, 2011) em meados de 2003 a sociedade já havia produzido e armazenado mais dados digitais em *pendrives*, HD's, CD's e memórias do que em meios analógicos como fotos, livros ou jornais. Em 2008 e 2009 foram produzidos cerca de 6,3 quintilhões de bytes todos os dias além de 90% dos dados do mundo foram criados nos últimos anos, isso devido a adesão das grandes empresas à internet, criação de redes sociais e dispositivos móveis. Em pouquíssimo tempo a humanidade passou a armazenar muitos dados, em 1986 tudo o que já se havia produzido ocupava 2600 petabytes, 20 anos depois, em 2007, ocupava-se 295 exabytes, e em apenas 4 anos depois já se havia inalgurado

uma nova unidade, em 2011 já se havia acumulado 1,6 zettabytes.

2 A *Big Data*

O avanço da tecnologia para armazenar dados tornou-se cada vez mais rápido devido ao aumento exponencial de geração de dados, através da internet e diversos dispositivos como celulares e computadores. Antigamente os *hardwares* para armazenar esses dados eram mais robustos, caros e não suportavam a quantidade de *bytes* que eram necessários para guardar. Agora, a sociedade produz mais informações que precisam ser guardados em sistema com um custo menor, maior capacidade e portáteis.

Devido a essa quantidade enorme de dados que não são estruturados e são gerados a cada segundo, deu-se o nome de *Big Data*. Atualmente essa tecnologia está se tornando essencial para aprimorar o conhecimento da sociedade, seja em tomar decisões inteligentes de projeto ou em negócios, como definir estratégias de *marketing* para aumentar a produtividade e reduzir os custos. Esses dados são originados a partir de redes sociais, *site* de internet, como também registros de compras, médicos, impressões digitais e entre outros.

Como já dizia (TAURION, 2013), no livro "*Big Data*", a aplicabilidade do *Big Data* está no tratamento desse volume de dados, que vem de variadas fontes e que demandam alta velocidade de processamento, na busca por um valor. Esse valor, que podem demonstrar diversos benefícios como a descoberta de padrões de um experimento, oscilações em ações da bolsa, e entre outros, que através dos dados podem divulgar características necessárias para ampliar o conhecimento da sociedade.

Para lidar com essa quantidade impactante de dados, e com a velocidade que são gerados, foi necessário desenvolver ferramentas para que o *Big Data* pudesse proporcionar informações relevantes para as empresas. Assim, foram definidos conceitos da estratégia do *Big Data* que explicam o impacto e as implicações desse tipo de tecnologia, chamados de 5'V. O livro de (RIJMENAM, 2014), explica os conceitos como:

- Volume: A quantidade de dados gerados atualmente é inimaginável, só em 2011 a sociedade já produziu mais de 1,8 *zettabytes* de dados;
- Variedade: Cerca de 90% dos dados não são estruturados e eles surgem de vários formatos. Cada tipo de dado necessita de análises e ferramentas diferentes para interpretar;
- Velocidade: É a agilidade em que os dados são criados, armazenados e analisados. Atualmente é necessário que haja uma análise nos dados em tempo real, devido à enorme quantidade de dados que são gerados;
- Veracidade: Dados obtidos errados podem causar problemas significantes para as empresas. É necessário que o dado obtido seja fielmente correto;
- Valor: O armazenamento dos dados podem fornecer um valor elevado de mercado para uma empresa, consumidores e para a sociedade. Como é o caso da indústria de saúde do estado de McKinsey, EUA, que possui um valor de \$300 bilhões em *Big Data*.



Figura 1 – Os 5V's da *Big Data*

Fonte (SEMRUSH,)

Atualmente esse valor atribuído à *Big Data*, para não só armazenar e transmitir os dados, como também processar eles, gera mais riquezas do que produzir soja, como é comentado no livro (FLORIDI, 2014). A quantidade de dados gerados em 2015 passaram os oito *zettabytes*, estima-se que em 2025 serão possíveis mais de dois bilhões de pessoas terem o binoma sequenciado que ultrapassam os *exabytes* de dados que serão gerados, como é dito no artigo (STEPHENS et al., 2015).

A variedade dos tipos de dados é extensa e pode-se listar elas como dados estruturados, semiestruturados, não estruturados e estruturados complexos. Os dados estruturados são aquelas informações armazenadas em banco de dados e possuem um formato restrito, como dados de sensores, da geografia local em tempo real e entre outros. E pode-se exemplificar os dados não estruturados como quaisquer documentos, arquivos, gráficos que não foram codificados, como documentos em *Word*, *emails* e entre outros.

2.1 Uso da tecnologia na sociedade

As informações obtidas através de dados gerados pela sociedade na internet, estão sendo cada vez mais utilizados por empresas para criar estratégias de *marketing* e estimular os clientes a aderir a negócios. As empresas usufruem do *Big Data* para satisfazer seus clientes, como é o caso dos serviços da *Netflix* e *Spotify*, que utilizam o histórico de dados do cliente para sugerir serviços que provavelmente o usuário irá aderir, isso acontece em diversas empresas como a *Amazon* que sugerem anúncios de compras baseado no histórico de pesquisas do cliente.

Numa matéria realizada nos Estados Unidos, em (VENTUREBEAT,), foi anunciado que a *Microsoft* fará uma parceria com a Universidade Federal de Minas Gerais para realizar pesquisas que poderão prever engarrafamentos com até uma hora de antecedência. De acordo com o site, "*The Traffic Prediction Project*", o projeto de previsão de trânsito, deverá usar dados com números históricos do departamento de transportes, câmeras das

estradas, mapas de trânsito do *Bing*, como também *posts* nas redes sociais do motoristas. Assim, com o auxílio da *Big Data* será possível facilitar a vida da sociedade, como já dizia Eric Siegel no livro (SIEGEL, 2013), os dados que nós coletamos atualmente, nos permitem ver as coisas que até pouco tempo atrás eram grandes demais para enxergamos.

Com essa abertura de divulgação de dados dos clientes, a empresa *Telefônica* utilizando análises de *Big Data*, deu suporte as ações do governo para combater um surto de gripe suína, ao obter dados da rede de celulares das pessoas para identificar como elas se locomovem durante o surto da doença.

Um caso de manipulação dos dados pode ser visto no artigo (MUSICUS; TAL; WANSINK, 2015), ao mostrar que os mercados sabiam aonde posicionar as compras, após testes com diferentes combinações, e foi visto que ao colocar cereais infantis, em prateleiras mais baixas, e cereais para adultos, nas prateleiras elevadas, estavam trazendo grandes resultados de vendas. Outro fator interessante que é mostrado no artigo, é o fato dos personagens ilustrados nas caixas de cereais estarem olhando nos olhos dos consumidores. Essa ferramenta permitiu aumentar em 16% a venda dos cereais, apenas ao entender como manipular os dados.

Uma reportagem realizada no *The New York Times Magazine*, por Charles Duhigg em (DUHIGG, 2012), mostra como uma loja de mercado americana da marca *Target*, descobriu que uma mulher estava grávida antes que o pai dela soubesse. Isso foi possível depois que o mestre em estatística Andrew Pole, treinou o sistema da loja com o padrão de consumo das clientes grávidas, sabendo assim que produto que era comprado pela gestantes, ao saber gerenciar os dados obtidos, foi possível descobrir a gravidez da garota.

2.2 Possíveis riscos à sociedade

Para muitos engenheiros a parte matemática era difícil de compreender e para os matemáticos faltava a visão de engenharia no que foi publicado por Claude Shannon em 1948, que aborda o que resta se a informação pudesse ser separada do seu conteúdo semântico. Em (SHANNON, 1948) trata a informação como incerteza, define ela como entropia, a medida de desordem na termodinâmica. A incerteza por sua vez, está ligada à previsibilidade do valor que será assumido por uma dada variável aleatória. A teoria da informação é baseada em probabilidade e estatística e define a informação como função logarítmica de base 2, o bit.

O crescimento do poder de computação é de forma exponencial, e junto com ele o que a sociedade vive é um estado de saturação de informações. Este estado cria uma ansiedade por informações, e o bombardeio de dados muitas vezes fracassa em transmitir o que realmente era necessário a se saber. Em (GLEICK, 2013) conta como em pouco tempo depois de surgir a teoria da informação de (SHANNON, 1948) surge uma nova síndrome, a "fadiga de informação", a qual se entra em um estado mental de exaustão pelo contato desmasiado de informações.

A massividade da informação faz com haja dificuldade do indivíduo escolher e retirar o que lhe é útil. A informação passa a ser facilmente manipulada, os dados gerados podem ser usados para aprender os comportamentos da sociedade e utilizados para controlar através de informações falsas misturadas a verdades. É possível influenciar a população em uma eleição de um presidente através de redes sociais, como *FaceBook*. A informação desmasiada prejudica o julgamento (BARTLETT; GREEN, 1966).

Na Índia foi feita uma pesquisa que demonstrou a influência resultado do buscador

de internet nas eleições. Uma amostra de 2150 eleitores indecisos teve, no dia das eleições, a ordem dos 30 primeiros resultados de suas pesquisas no Google relativas às eleições alterada para favorecer um ou outro candidato. Essa simples mudança na ordem de apresentação dos resultados foi capaz de influenciar mais de 20% dos indecisos em favor do candidato beneficiado pelo buscador. O resultado mostra um ganho de 2% do eleitorado total e como cerca de um quarto das eleições ao redor do mundo são ganhas por uma margem inferior a 3%, essa manipulação possa impactar, ou, já esteja impactando os resultados de votações em diversos países (EPSTEIN; ROBERTSON, 2015).

Uma tendência contemporânea das grandes empresas de internet é trabalhar no aprimoramento da capacidade dos algoritmos de entregar mensagens cada vez mais personalizadas para seus usuários, o que leva a uma perigosa consequência; a de ser preso em uma espécie de “bolha” onde não se é expostos a informações que desafiam o pensamento e que expandam a forma de enxergar o mundo. (PARISER, 2011)

Além dos prejuízos relacionados a saúde mental, manipulação de massas e geração de "bolhas", deve-se levar em consideração que a *Big Data* é uma importante ferramenta mas que a análise de significado depende do contexto. Existem diversas formas de expressão que podem confundir a avaliação de uma máquina, como por exemplo, uma frase ambígua ou irônica. Assim como o cérebro humano tem dificuldades e fica confuso, a inteligência artificial também ficará fazendo correlações erradas o que poder trazer prejuízos, como é o caso de *Taytweets* o robô de inteligência artificial da *Microsoft* que se tornou racista em menos de um dia com interação com humanos em redes sociais.

3 Engenharia e a ciência de dados

Como foi comentando anteriormente, como a sociedade está cada vez mais produzindo dados e o poder computacional para armazenar essas informações também está aumentando, torna-se possível analisar mais dados e extrair lições, conhecimentos, para isso é utilizado a ciência de dados.

Lidar com os dados é uma tarefa complicada e requer extremo cuidado, pode-se exemplificar um caso do matemático húngaro Abraham Wald, em (MANGEL; SAMANIEGO, 1984). Que na segunda guerra mundial, recebeu uma tarefa de otimizar a blindagem de aviões em combate ingleses, o problema que não era possível blindar todo avião por ser pesado demais, e o avião não conseguiria voar, assim era necessário planejar onde colocar a parte blindada. Então, a equipe de pesquisa de estatística do matemático checou os aviões que voltavam do combate e anotavam aonde eles tinham furo de bala, e com isso viam as partes do avião que mais levavam tiros. Sendo assim, essa seria a região para colocar a blindagem. No entanto, Abraham Wald identificou que na realidade essas não eram as áreas que mais levavam tiros, mas sim que os aviões que voltavam e podiam ser observados eram os aviões sobreviventes. Assim, através da análise de Wald, pode-se analisar que os aviões que eram acertados no motor não sobreviviam. Abraham soube lidar com os dados, sendo dito como um cientista de dados, e trazer benefícios para a engenharia aeronáutica.

O conceito de *internet* das coisas, que em outras palavras é uma rede de objetos físicos (sensores, veículos,...) capaz de coletar e transmitir dados, trouxe ainda mais necessidade de armazenar dados. Máquinas tecnológicas em indústrias utilizam cada vez mais essa conexão com a *internet*, para obter dados de processos e operar automatizado.

O artigo (BOLLEN; MAO; ZENG, 2011) mostra como as informações obtidas pelos usuários de uma rede social chamada *Twitter*, junto com análises de texto, pra saber

se estavam elogiando ou não uma empresa, pudesse prever que as ações de uma certa empresa estivessem em alta ou em baixa. Sendo que muitos estão usando esses dados para planejar investimentos nas ações.

Uma das tecnologias vigentes é o aprendizado de máquinas (*machine learning*), trata-se de uma nova ferramenta que ajuda a achar as respostas de um problema através de dados. O aprendizado de máquinas deixa o algoritmo checar os dados, aprender como os dados estão distribuídos e faz com que a máquina construa regras para prever como os dados vão se comportar. Mesmo o programador não sabendo que regras são essas. É necessário que haja um padrão de dados para o aprendizado de máquinas, e que o cientista dos dados garanta que os dados estão limpos, como por exemplo, o ponto e a vírgula nos números são usados de maneira consistente. Sendo assim, uma análise normal de dados ruins geram resultados ruins, se implementado no aprendizado de máquinas, os dados ruins geram modelos errados e não tem dado novo que possa corrigir o modelo errado. O aprendizado de máquinas é usar dados e testes que possam refinar como o programa ou computador podem fazer a análise, assim usando um mundo de dados coletados pela ciência de dados.

Com o armazenamento de dados em paralelo ao aprendizado das máquinas, fica o questionamento da substituição do operário pela máquina. No livro ([BRYNJOLFSSON; MCAFEE](#),), os autores dizem que nos Estados Unidos, durante a troca do campo para as fábricas, a produtividade e os salários aumentaram juntos até a década de 90. De lá para cá, a produtividade continua aumentando, mas enquanto o que uma família de classe média ganha não. De acordo com os autores, isso deve-se à automação, visto que o operário em 2013 produzia o dobro do que alguém em 1973, mas não ganhava mais do que 10%. No livro ([FORD, 2015](#)), o autor cita que a fase onde o desenvolvimento econômico trazia novos postos de trabalho e aumento de poder de consumo para todo mundo, pode estar acabando.

Considerações finais

O dilúvio de dados já é uma realidade da sociedade, a quantidade de informações que são compartilhadas pela *internet* possibilitam romper fronteiras do conhecimento. Basta agora, reconhecer qual é o papel do engenheiro e do cientista de dados para manipular essa quantidade alarmante de dados, a *Big Data*, e trazer benefícios para as pessoas, como o histórico nos trás de informação. Sabe-se que no passado, o significado da escrita pôde trazer pensamentos lógicos mais complexos e armazenar conhecimentos para a sociedade. Atualmente com o avanço tecnológico, a quantidade de dados que são gerados é impactante, toda a história da humanidade pode ser guardada em armazenamentos digitais desde 1986.

O *Big Data* tornou-se uma ferramenta muito poderosa e importante tanto para criar estratégias de mercado, como identificar preferências de usuários de serviços. É de extrema importância saber lidar com a aquisição de dados e saber como manipular eles, a velocidade, variedade e o valor que os dados são entregues se transformaram em fatores que necessitam ser visto em tempo real, pois eles estão sendo constantemente gerados. Assim, essa fragilidade da sociedade que está exposta a compartilhar informações, dados para a *internet*, fazem com que empresas multinacionais usufruem dessa *Big Data* para satisfazer seus clientes, criando estratégias de mercado. Portanto, tem-se que analisar não só a área comercial, mas também os benefícios dos dados para a sociedade, como criação de novas tecnologias, prevenção de risco, cuidados com a saúde, que mais do que dados

estatísticos, são dados que precisam ser lidados para progredir o conhecimento.

No entanto, sabe-se que devido a essa quantidade enorme de dados dispostos na *internet*, há uma saturação de informação. Complicando a sociedade em conseguir identificar notícias, informações falsas, criando também um estado de exaustão mental pelo contato demorado de informação. Empresas e governos que possuem um maior contato com o *Big Data* podem interferir decisivamente em rumos de uma sociedade, muitas vezes seguindo ideologias equivocadas, onde o único privilegiado é o estado e seus interesses nacionais.

Portanto, é inevitável entender, analisar e posicionar o papel do engenheiro na sociedade com tamanha disposição de dados que podem ser manipulados. Tecnologias surpreendentes que antigamente eram vista apenas em filmes de ficção, estão se tornando cada vez mais real, como o *machine learning*. Com isso, deve-se pensar mais em criatividade e trabalho humanizado, para que a sociedade possa participar da construção histórica de uma nação mais forte e soberana.

Referências

BARTLETT, C. J.; GREEN, C. G. Clinical prediction: Does one sometimes know too much? *Journal of Counseling Psychology*, Wm. C. Brown Co., v. 13, n. 3, p. 267, 1966. Citado na página 5.

BOLLEN, J.; MAO, H.; ZENG, X. Twitter mood predicts the stock market. *Journal of computational science*, Elsevier, v. 2, n. 1, p. 1–8, 2011. Citado na página 6.

BRYNJOLFSSON, E.; MCAFEE, A. *A Segunda Era Das Máquinas*. ELSEVIER/ALTA BOOKS. ISBN 9788576089148. Disponível em: <<https://books.google.com.br/books?id=O5NLtAEACAAJ>>. Citado na página 7.

DUHIGG, C. How companies learn your secrets. *The New York Times*, The New York Times Company Nova York, v. 16, p. 2012, 2012. Citado na página 5.

EPSTEIN, R.; ROBERTSON, R. E. The search engine manipulation effect (seme) and its possible impact on the outcomes of elections. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 112, n. 33, p. E4512–E4521, 2015. Citado na página 6.

FLORIDI, L. *The fourth revolution: How the infosphere is reshaping human reality*. [S.l.]: OUP Oxford, 2014. Citado na página 4.

FORD, M. *Rise of the Robots: Technology and the Threat of a Jobless Future*. [S.l.]: Basic Books, 2015. Citado na página 7.

GLEICK, J. *A informação: uma história, uma teoria, uma enxurrada*. [S.l.]: Editora Companhia das Letras, 2013. Citado 2 vezes nas páginas 2 e 5.

HILBERT, M.; LÓPEZ, P. The world's technological capacity to store, communicate, and compute information. *science*, American Association for the Advancement of Science, v. 332, n. 6025, p. 60–65, 2011. Citado 2 vezes nas páginas 1 e 2.

MANGEL, M.; SAMANIEGO, F. J. Abraham wald's work on aircraft survivability. *Journal of the American Statistical Association*, Taylor & Francis Group, v. 79, n. 386, p. 259–267, 1984. Citado na página 6.

MUSICUS, A.; TAL, A.; WANSINK, B. Eyes in the aisles: why is cap'n crunch looking down at my child? *Environment and Behavior*, SAGE Publications Sage CA: Los Angeles, CA, v. 47, n. 7, p. 715–733, 2015. Citado na página 5.

PARISER, E. *The filter bubble: What the Internet is hiding from you*. [S.l.]: Penguin UK, 2011. Citado na página 6.

RIJMENAM, M. V. *Think bigger: Developing a successful big data strategy for your business*. [S.l.]: Amacom, 2014. Citado na página 3.

SEMRUSH. *Big data: conheça os 5 V's e sua aplicação prática para PMEs*. Disponível em: <<https://pt.semrush.com/blog/big-data-conheca-os-5-vs-e-sua-aplicacao-pratica-para-pmes/>>. Citado na página 4.

SHANNON, C. E. A mathematical theory of communication. *Bell system technical journal*, Wiley Online Library, v. 27, n. 3, p. 379–423, 1948. Citado na página 5.

SIEGEL, E. *Predictive analytics: The power to predict who will click, buy, lie, or die*. [S.l.]: John Wiley & Sons, 2013. Citado na página 5.

STEPHENS, Z. D. et al. Big data: astronomical or genetical? *PLoS biology*, Public Library of Science, v. 13, n. 7, p. e1002195, 2015. Citado na página 4.

TAURION, C. *Big data*. [S.l.]: Brasport, 2013. Citado na página 3.

TURING, A. M. On computable numbers, with an application to the entscheidungsproblem. a correction. *Proceedings of the London Mathematical Society*, Wiley Online Library, v. 2, n. 1, p. 544–546, 1938. Citado na página 2.

VENTUREBEAT. *How Microsoft's using big data to predict traffic jams up to an hour in advance*. Disponível em: <<https://venturebeat.com/2015/04/03/how-microsofts-using-big-data-to-predict-traffic-jams-up-to-an-hour-in-advance/>>. Citado na página 4.