



PERFIL DE BIG DATA: tendências profissionais e tecnológicas

BIG DATA PROFILE: professional and technological trends

Luciano Bérغامo^I
Sabrina dos Santos Teixeira^{II}

RESUMO

O profissional de Big Data, engenheiro, é responsável por automatizar os processos de grande quantidades de dados e tratá-los de forma correta para tomadas de decisões e coloca em função áreas de programação que ajuda no cruzamento dos dados, limpeza e geração de resultados. O contexto Big Data engloba muitos setores da área de TI, tornando hoje um dos segmentos da Tecnologia da Informação em grande crescimento e sempre buscando profissionais de alto nível. Além de Cientistas de Dados e Engenheiros de Dados existem os Engenheiros de Machine Learning, que são profissionais responsáveis pelo aprendizado das máquinas. O objetivo desta pesquisa é identificar profissões emergentes relacionadas ao contexto da Ciência de Dados de forma a oferecer aos públicos, acadêmico e empresarial, alternativas de formação e capacitação nesta área e apresentar algumas obras e autores relacionados ao assunto. Os procedimentos metodológicos estão associados a uma pesquisa bibliográfica com a finalidade de identificar as profissões emergentes no mercado de Tecnologia da Informação voltado para a valorização dos dados. Entre as profissões pesquisadas e discutidas, nota-se que Cientistas e Engenheiros de Dados têm funções similares, porém, o Engenheiro de Dados é mais envolvido com o estudo da estrutura de suporte aos dados, incluindo *hardware* e *software*. Já o Cientista de Dados trabalha no quesito de linguagem de máquina “*Machine Learning*”, onde os processos são automatizados para filtrar grandes volumes de dados e obter os melhores resultados possíveis, também denominados como dados limpos.

Palavras-chave: Big Data, Machine Learning, Cientista de Dados, Formação

ABSTRACT

The Big Data professional, an engineer, is responsible for automating the processes of large data and treating them correctly for decision making and puts programming areas in function that helps in the crossing of data, cleaning and generation of Results. The Big Data context encompasses many sectors of the IT area, today making one of the segments of Information Technology in great growth and always seeking high-level professionals. In addition to Data Scientists and Data Engineers there are Machine Learning Engineers, who are professionals responsible for machine learning. The objective of this research is to identify emerging professions related to the context of Data Science in order to offer audiences, academic and business, training and training alternatives in this area and present some works and authors related to the subject. Methodological procedures are associated with bibliographic research

^I Prof. Me. da Faculdade de Tecnologia (FATEC) de Mococa – São Paulo – Brasil. E-mail: lbpsolucoes@gmail.com

^{II} Estudante do Curso Superior de Tecnologia em Gestão da Tecnologia da Informação da Faculdade de Tecnologia (FATEC) de Mococa – São Paulo – Brasil. E-mail: sabrinasantos1798@hotmail.com



in order to identify emerging professions in the Information Technology market aimed at valuing data. Among the professions researched and discussed, it is noted that Scientists and Data Engineers have similar functions, however, the Data Engineer is more involved with the study of the data support structure, including hardware and software. Data Scientist, on the other hand, works on the machine language ite "Machine Learning", where processes are automated to filter large volumes of data and get the best possible results, also called clean data.

Keywords: Big Data, Machine Learning, data scientist, Training

Data de submissão do artigo: 04/06/2019.

Data de aprovação do artigo: 01/10/2019.

DOI:

1 INTRODUÇÃO

A necessidade de administrar dados tornou-se uma das maiores preocupações das organizações atuais. Ao mesmo tempo que esta evolução tecnológica oferece benefícios a sociedade, traz consigo grandes obstáculos. Um problema que se destaca neste contexto é a falta de profissionais capacitados para fornecer suporte tecnológico de modo que as empresas consigam extrair os dados, compila-los e transformá-los em informações que darão apoio a tomada de decisão.

Diante deste cenário, pergunta-se: Quais são as profissões que habilitam pessoas para lidar com esta nova tendência tecnológica?

O objetivo desta pesquisa é identificar profissões emergentes relacionadas ao contexto da Ciência de Dados de forma a oferecer aos públicos, acadêmico e empresarial, alternativas de formação e capacitação nesta área e apresentar algumas obras e autores relacionados ao assunto.

A pesquisa se justifica por tratar de um tema atual que vai de encontro ao pensamento contemporâneo tradicional, quebrando paradigmas referentes a análise e interpretação dos dados. Justifica-se também pela percepção da escassez de profissionais qualificados nesta área. A pesquisa se enquadra ao tema central do evento (Revolução Digital: Tendências Tecnológicas no Mundo Moderno).

2 SISTEMAS DE APOIO A DECISÃO

Em uma organização é fundamental que os dados sejam valorizados de tal maneira que seja possível preparar cenários que forneçam informações que oferecerão estratégias para decisões sobre investimentos. Para isso, são implantados os sistemas de apoio a decisão.

Segundo O'Brien (2002), um sistema de informações gerenciais gera informações que subsidiam os administradores nas tomadas de decisão. Deste modo, tais sistemas auxiliam os administradores na busca de soluções que propiciem às organizações melhoria contínua de seus processos, produtos e serviços, pois levam aos gestores informações que o decisório demanda.



2.1 Business Intelligence

Segundo Bergamo (2015) é possível representar o conceito de *Business Intelligence (BI)* por meio de definições gerais e considerações específicas.

Definições Gerais:

- As empresas estão acumulando dados processados pelos sistemas de informação, visando apoio à reflexão dos resultados para melhoria dos negócios. Entre os diversos tipos de Sistemas de Informação, destacam-se os Sistemas de Apoio a Tomada de Decisão.
- Na indústria contemporânea, existe a crescente necessidade do processamento dos dados em tempo real, uma vez que pode haver uma grande massa de dados gerados na operação, sendo necessária a implantação de sistemas de informação para a gestão e cruzamento dos dados com vista a melhorar a produção.
- Por meio do uso de sistemas de apoio a decisão detectou-se que, nas sextas-feiras à tarde, os homens norte-americanos que compravam fraldas no *Wal-Mart* tinham a tendência de também comprarem cervejas. As vendas aumentaram quando os encarregados colocaram as cervejas ao lado das fraldas.
- Os sistemas de informação convencionais geralmente processam e imprimem informações em relatórios triviais como: O boletim de um aluno ou uma lista de cargos e salários dos funcionários.

Considerações Específicas

- Entre os sistemas de apoio a tomada de decisão, existem os Sistemas de Informações Inteligentes, mais conhecidos como *BI*. Sistemas de *BI* geram relatórios diferentes dos sistemas convencionais e impulsionam as organizações a se adaptarem às mudanças do negócio.
- *BI* não é um sistema comum, ele fornece recursos para o cruzamento de dados do negócio, gerando um formato de consulta dinâmica sobre os dados processados.
- É um sistema que recebe dados de várias fontes e transforma-os em informações consolidadas.
- A arquitetura de *BI* consiste fundamentalmente de um armazém de dados e a partir dele oferece uma variedade de consultas, relatórios e ferramentas analíticas.

2.2 Inteligência Artificial

A *inteligência artificial (IA)* é a próxima onda de inovação, impulsionada pelos avanços em capacidade computacional, pela capacidade de armazenar grandes volumes de dados na nuvem a um custo mínimo e pela facilidade de acesso a algoritmos avançados.

A *IA* é representada por termos como *Machine Learning*, *Deep Learning*, processamento natural de linguagem e análises preditivas.

É importante destacar que por trás da *Inteligência Artificial* há profissionais importantes como o Cientista de *Machine Learning* e o Engenheiro de *Machine Learning*.

2.2.1 Machine Learning

Segundo Matos (2018) *Machine Learning* é um conjunto de regras e procedimentos, que permite que os computadores possam agir e tomar decisões baseados em dados ao invés



de ser explicitamente programados para realizar uma determinada tarefa. Programas de Machine Learning também são projetados para aprender e melhorar ao longo do tempo quando expostos a novos dados.

O aprendizado de máquina é um subcampo da *inteligência artificial (IA)*. O objetivo do aprendizado de máquina geralmente é entender a estrutura dos dados e adequar esses dados a modelos que possam ser compreendidos e utilizados pelas pessoas (TAGLIAFERRI, 2017).

2.3 Ciências de dados

Geralmente, a ciência de dados é a aplicação de métodos quantitativos e qualitativos para resolver problemas relevantes e prever resultados. Uma das principais revelações de hoje, com a vasta e crescente quantidade de dados, é que o conhecimento e a análise dos dados não podem caminhar separados.

Uma pesquisa norte americana chamada *Data Science Predictive Analytics*, traduzida em nossa língua como Análise Preditiva da Ciência de Dados, apresenta que os estudantes têm dificuldades para compreender o pensamento analítico em análise de dados (DUMBILL; LIDDY; MUELLER, 2013).

De acordo com Dumbill, Liddy e Mueller (2013) é mais fácil ensinar linguagem de programação para um aluno da área da ciência da computação do que ensinar o domínio de conhecimento para aplicação de análises quantitativa e qualitativa sobre os dados.

Ainda nesta pesquisa, Dumbill, Liddy e Mueller (2013) cita um texto de Shelly Farnham, pesquisador da Microsoft sobre a preocupação em contratar pessoas com domínio em análise de dados:

Quando estamos procurando pessoas, estagiários ou pessoas em tempo integral em nossa equipe, definitivamente procuramos pessoas que tenham experiência na análise de dados. (FARNHAM *apud* DUMBILL; LIDDY; MUELLER, 2013, p. 25).

Assim, tanto acadêmicos quanto profissionais de mercado devem ter habilidades analíticas, entender do negócio e conhecer os conceitos de gestão. Como Provost e Fawcett (2013, p. 52) escrevem: “A ciência de dados envolve muito mais do que apenas algoritmos de mineração de dados. Cientistas de dados bem-sucedidos devem ser capazes de visualizar problemas de negócios a partir de uma perspectiva de dados”

De acordo com Provost e Fawcett (2013), avanços em tecnologia de computação e exploração de dados acontecem em nossa época por questão do grande volume de dados disponíveis para análises. Do grande volume de dados, nasce o conceito de *Big Data*.

Cukier e Mayer (2013) apontam que *Big Data* é um novo conceito baseado em dados, oferecendo o conceito de pensar sobre quais dados integrados para resolver questões de negócios, e problemas que não eram resolvidos anteriormente.

Provost e Fawcett (2013) afirmaram que a ciência de dados envolve princípios, processos e técnicas para entender fenômenos através da análise (automatizada) de dados. Isto indica que, por meio da ciência de dados é possível obter respostas sobre análise em grande escala de dados como: opiniões de usuários em redes sociais, respostas sobre dados de acessos a vídeos, filmes, áudios e outras mídias pesquisadas ou acessadas na internet.

Inditica-se como os três pilares da ciência de dados: dados, tecnologias e pessoas.

Dados: referem-se as áreas de domínio, como dados relacionais, dados não relacionais



(por exemplo, dados não estruturados e semiestruturados, como dados de mídia social e dados da web) e dados do sensor;

Tecnologias: inclui ecossistemas como *Hadoop*, *Bancos de Dados NoSQL*, mineração de dados, aprendizado de máquina e computação em nuvem;

Pessoas: Incluem cientistas da computação, estatísticos, especialistas de domínio, cientistas de dados e analistas de negócios

A ciência de dados engloba *Big Data* como um campo multidisciplinar que precisa de análise iterativa com especialistas de domínio. A ciência de dados também tem um forte lado humano.

Por exemplo: design de interface do usuário e resultados apresentados por meio de gráficos e relatórios resumidos. Entre os três pilares apresentados por o mais importante são as pessoas.

“É possível comprar mais computadores, armazenamentos e ferramentas para processar efetivamente grandes volumes de dados, mas a capacidade humana não aumenta; educar as pessoas, chamado cientistas de dados, é fundamental para enfrentar os desafios da era do *Big Data*” (EXPERT, 2016 p. 364-373).

2.4 Big Data

Big Data pode ser entendido como o fenômeno que se refere a um volume de dados que é difícil de tratar através de tecnologias clássicas de banco de dados (HASHEM *et al.*, 2015).

Outra definição amplamente adotada é caracterizar o fenômeno sob alguns aspectos, os conhecidos “V”s, que podem ser 3, 4 ou 6, dependendo dos estudos. Por exemplo, um dos estudos que adota os 4 “V”s os definem como: volume, variedade, velocidade e valor (GANTZ; REINSEL, 2011). Considerando-se os 4 “V”s, o conceito de *Big Data* aproxima-se do contexto de Inteligência Analítica. A análise de dados que se enquadram no conceito de *Big Data* são frequentemente chamadas de Big Data Analytics, e há autores que garantem ser o mesmo que Inteligência Analítica (CHEN; CHIANG; STOREY, 2012), principalmente no que tange as tecnologias utilizadas para lidar com dados com tais características. *Big Data* não tem valor para tomada de decisão, a menos que seja realizada uma análise dos dados (TIEN, 2013) para ganho de inteligência. Esse é o elo que une *Big Data* e Inteligência Analítica.

As plataformas atuais de *Big Data* são suportadas por várias ferramentas analíticas de processamento, bem como visualização dinâmica. Essas plataformas permitem extrair conhecimento e valor do ambiente dinâmico complexo. Eles também apoiam a tomada de decisões através de recomendações e detecção automática de anomalias, comportamento anormal ou novas tendências.

Portanto, *Big Data* se apoia nos princípios do BI (*Business Intelligence*) que utilizará seus dados para fornecer informações ágeis para a tomada de decisão.

3 PROCEDIMENTOS METODOLÓGICOS

Os procedimentos metodológicos estão associados a uma pesquisa bibliográfica com a finalidade de identificar as profissões emergentes no mercado de Tecnologia da Informação voltado para a valorização dos dados. Para escrever o artigo foram utilizados como referência obras e artigos acadêmicos.



4 RESULTADOS E DISCUSSÕES

Os resultados dessa pesquisa relacionam-se às profissões e obras relacionadas ao big data. Muitas profissões estão surgindo neste novo milênio, para Big Data é possível destacar novas funções como: Engenheiro de Dados, Engenheiro de Big Data, Arquiteto de Soluções de Big Data, Cientista e Engenheiro de Machine Learning, Especialista em Business Analytics (analista de negócios), Desenvolvedor de Visualização de Dados, Estatístico e Cientista de Dados.

4.1 Engenheiro de dados

O Engenheiro de Dados é o responsável por garantir que os dados estarão disponíveis para a análise de forma segura. Engenheiros de dados constroem enormes reservatórios para Big Data. Eles desenvolvem, constroem, testam e mantêm arquiteturas, tais como bancos de dados e sistemas de processamento de dados em grande escala.

Uma vez que estes imensos reservatórios de dados estejam criados, Cientistas de Dados podem aplicar suas técnicas analíticas e extrair informação relevante. O trabalho do Engenheiro de Dados está diretamente ligado a infraestrutura de TI.

4.2 Engenheiro de big data

Esta carreira é uma extensão da carreira de Engenheiro de Dados, mas, com foco em Big Data. Montar uma infraestrutura para armazenamento e processamento de grandes conjuntos de dados não é tarefa fácil. Hadoop, Spark, Cassandra, Hive, Hbase, Pig, Sqoop, MongoDB, API de integração.

Cada solução será apropriada para um projeto específico e cabe a este profissional provisionar o Big Data de modo que os analistas e cientistas de dados possam aplicar seus modelos preditivos e resolver problemas de negócio. Assim como o Engenheiro de Dados, o Engenheiro de Big Data tem seu trabalho diretamente ligado à infraestrutura de TI.

4.3 Arquiteto de soluções de big data

Arquitetos de dados criam projetos para sistemas de gestão de dados. Depois de avaliar potenciais fontes de dados da empresa (interna e externa), os arquitetos projetam um plano para integrar, centralizar, proteger e manter esses dados. Isso permite que os funcionários acessem informações críticas no lugar certo e na hora certa. A arquitetura da solução de Big Data projetada por este profissional, será implementada pelo Engenheiro de Dados ou Engenheiro de Big Data.

4.4 Cientista de *Machine Learning*

Este profissional trabalha com pesquisa e desenvolvimento de algoritmos que são usados para criar sistemas inteligentes. Eles constroem sistemas para recomendação de produtos ou para prever demandas em determinados produtos ou serviços e exploram Big



Data para extrair padrões dos dados. Se você gosta de pesquisa e possui forte conhecimento em Matemática e Estatística, considere esta carreira como opção.

4.5 Engenheiro de *Machine Learning*

O trabalho do Engenheiro de Machine Learning é parecido com o da carreira anterior, mas seu foco está em criar uma solução de software que permita resolver um problema de negócio através de modelos preditivos. Enquanto o Cientista de Machine Learning tem como objetivo pesquisar e desenvolver novos algoritmos, o Engenheiro de Machine Learning tem como objetivo aplicar esses algoritmos e criar soluções. Aplicar algoritmos de Machine Learning implica em conhecimento de Matemática, Estatística, procedimentos de limpeza e pré-processamento de dados e pelo menos uma linguagem ligada a Data Science, como R ou Python.

4.6 Especialista em *Business Analytics* (Analista de Negócios)

No atual complexo ambiente de negócios de uma organização, adaptabilidade, agilidade e capacidade de gerir mudanças constantes através de inovação, podem ser a chave para o sucesso. Os métodos tradicionais estão defasados no que diz respeito a conduzir à consecução dos objetivos quando as condições econômicas são desfavoráveis. É aí que entra a análise de negócios.

As corporações alcançam metas por meio de projetos que traduzem as necessidades dos clientes em novos produtos, serviços e geração de lucro. Os analistas de negócios podem fazer tudo acontecer de forma mais eficiente e eficaz. O principal objetivo do analista de negócios é ajudar as empresas a implementar soluções de tecnologia de uma forma eficaz em termos de custo, e assim, determinar os requisitos de um projeto ou programa e comunicá-los claramente aos interessados, facilitadores e parceiros.

4.7 Desenvolvedor de visualização de dados

O trabalho de Data Storytelling está se tornando crucial em qualquer projeto de Big Data. A habilidade de converter em um único gráfico ou Dashboard o que Petabytes de dados estão dizendo é quase uma arte e são muitas as técnicas e ferramentas disponíveis para este fim. Embora existam ferramentas que automatizem este processo (ou que pelo menos prometem automatizar), ainda assim esse profissional precisa conhecer mais do que tecnologia, mas conhecer quando usar o gráfico certo, como utilizar tabelas e relatórios e principalmente como usar a narrativa para apresentar a um público não técnico as conclusões de um projeto de análise de dados.

4.8 Gerente de *Analytics*

O Gerente de Analytics é o responsável pela equipe de análise de dados. Ele coordena o design, a configuração e a implementação de soluções de análise de dados, desde a infraestrutura, até a definição de ferramentas e processos de análise de dados. Fazer a gestão de projetos de Big Data não é tarefa simples e o gestor precisa ter habilidades de liderança e conhecimento técnico para compreender os desafios inerentes ao Big Data.



4.9 Estatístico

Embora a profissão de Estatístico não seja nova, está sendo reinventada pelo grande volume de dados e pelas novas ferramentas e soluções ligadas ao Big Data. A função deste profissional é aplicar técnicas estatísticas para a compreensão dos dados e ajudar as empresas a identificar tendências, fazer previsões e tomar decisões baseadas em dados.

Os estatísticos aplicam as teorias e métodos estatísticos para coletar, analisar e interpretar os dados. Eles trabalham para empresas envolvidas em pesquisa de mercado e opinião pública, para as indústrias que necessitam realizar o controle de qualidade e desenvolvimento de novos produtos, e – com frequência – para governos municipais, estaduais e federais.

4.10 Cientista de dados

Cientistas de dados são os grandes mineradores de dados. Eles recebem uma enorme massa de dados desorganizados (estruturados, semi-estruturados ou não-estruturados) e usam suas habilidades em matemática, estatística e programação para limpar, tratar, transformar e organizar esses dados. Em seguida, eles aplicam suas capacidades analíticas – conhecimento de negócio, compreensão contextual, ceticismo de suposições existentes e algoritmos de Machine Learning

5 CONSIDERAÇÕES FINAIS

Esta pesquisa proporcionou a possibilidade de identificar profissões emergentes relacionadas ao contexto da Ciência de Dados de forma a oferecer aos públicos, acadêmico e empresarial, alternativas para formação e capacitação nesta área.

Os objetivos foram alcançados, pois, com a pesquisa bibliográfica obteve-se a percepção da necessidade real existente nas organizações de encontrar profissionais capacitados para suprirem a demanda de serviços especializados em análise de dados.

Entre as profissões pesquisadas e discutidas, nota-se que Cientistas e Engenheiros de Dados trabalham têm funções similares, porém, o Engenheiro de Dados é mais envolvido com o estudo da estrutura de suporte aos dados, incluindo *hardware* e *software*. Já o Cientista de Dados trabalha no quesito de linguagem de máquina “*Machine Learning*”, onde os processos são automatizados para filtrar grandes volumes de dados e obter os melhores resultados possíveis, também denominados como dados limpos.

As pessoas com formação nas áreas de ciência da computação, estatística e matemática têm maior propensão a se tornarem cientistas de dados, desde que possuam o conhecimento e a perícia necessários para trabalhar em algumas das profissões apresentadas na sessão profissões e obras relacionadas ao *Big Data*. Já existem no Brasil programas formais de graduação em ciência de dados, bem como, *MBAs*, mestrados, especializações, entre outros ofertados por algumas universidades. Esses programas contribuirão com a formação dos profissionais do futuro.

REFERÊNCIAS

BÉRGAMO, L. **Aplicação de Business Intelligence como sistema de apoio à decisão para o setor editorial de ensino à distância (EAD)**. 2015. 126f. Dissertação (Mestrado em



Engenharia de Produção). Centro Universitário de Araraquara, Araraquara, 2015. Disponível em <https://www.uniara.com.br/arquivos/file/cursos/mestrado/engenhariaproducao/dissertacoes/2015/luciano-bergamo.pdf>. Acesso em: 25 fev. 2019

CHEN, H., CHIANG, R. e STOREY, V. **Business Intelligence and Analytics: do Big Data a Grande Impacto**. 2012. MIS Quarterly 36 (4): 1165–88.

CUKIER, Kenneth; MAYER, Viktor . **Big data: uma revolução que vai transformar a forma como vivemos, trabalhamos e pensamos**. 2013.

DUMBILL, E., LIDDY, E. D., J., MUELLER, K., & FARNHAM, S. **Educating the next generation of Data scientists**. Big Data, 2013.

EXPERT. **Systems How to Exploit the Value of Data in Today’s Big Data World**. 2016. Disponível em: <https://www.expertsystem.com/big-data/>. Acesso em: 04 jul. 2019

GANTZ, J., & REINSEL, D. **Extracting value from chaos**. 2011. IDC iView, 1142, 1-12.

HASHEM, I. A. T., YAQOUB, I., Anuar, N. B., MOKHTAR S., GANI, A., & Khan, S. U. **The rise of “Big Data” on cloud computing: Review and open research issues**. Information Systems. 2015.

MATOS, David. **Conceitos Fundamentais de Machine Learning**. 2018. Disponível em: <http://www.cienciaedados.com/conceitos-fundamentais-de-machine-learning/>. Acesso em: 04 jul. 2019.

O’ BRIEN, James A. **Sistemas de informação e as decisões gerenciais na era da internet**. São Paulo: Saraiva, 2002. 437p.

PROVOST, F. e FAWCETT, T. **Ciência de dados e sua relação com o Big Data e a tomada de decisões baseada em dados**. Big Data. 2013.

TAGLIAFERRI, Lisa. **An Introduction to Machine Learning**. 2017. Disponível em: <https://www.digitalocean.com/community/tutorials/an-introduction-to-machine-learning>. Acesso em: 04 jul. 2019.

TIEN, J. M. **Big Data: Unleashing information**. Journal of Systems Science and Systems Engineering. 2013.