



## De Business Intelligence a Data Science: um estudo comparativo entre áreas de conhecimento relacionadas

Alexandre de Oliveira Paixão<sup>1</sup>, Verônica Aguiar da Silva<sup>2</sup>, Asterio Tanaka<sup>1</sup>

<sup>1</sup> Programa de Pós-Graduação em Informática – Centro de Ciências Exatas e Tecnologia – Universidade Federal do Estado do Rio de Janeiro (UNIRIO)

<sup>2</sup> Coordenação de Informática – Instituto Federal Fluminense (IFF) – Campos de Goytacazes, RJ

[alexandre.paixao@uniriotec.br](mailto:alexandre.paixao@uniriotec.br), [v\\_aguiar@iff.edu.br](mailto:v_aguiar@iff.edu.br), [tanaka@uniriotec.br](mailto:tanaka@uniriotec.br)

### **Resumo**

*Ao longo dos anos, as empresas e organizações em geral têm se deparado com a contínua necessidade de tomar decisões em espaços de tempo cada vez menores. Ao mesmo tempo, cresce também a quantidade de dados disponíveis, fato que traz grandes desafios na coleta, armazenamento e processamento de tais dados, assim como em obter dos mesmos a informação e o conhecimento necessários para gerar a Inteligência de Negócio (do Inglês, Business Intelligence), permitindo, assim, o devido suporte na tomada de decisão. Recentemente, uma nova denominação tem sido usada para o tratamento de dados em diversas áreas do conhecimento, conhecida como Ciência de Dados (do Inglês, Data Science). O presente artigo tem como objetivo fazer um estudo comparativo dos diversos conceitos e técnicas usados em Inteligência de Negócio e Ciência de Dados, bem como de suas aplicações, traçando um paralelo entre esses dois termos contemporâneos.*

### **1. Introdução**

A economia baseada em conhecimento (Godin, 2006) trouxe novos desafios ao ambiente de negócios, desde o final do século passado, fazendo com que a aquisição, a análise e o acesso aos dados de uma organização, com o objetivo de suportar a tomada de decisão, se tornassem tarefas indispensáveis para o seu sucesso e, até mesmo, sobrevivência.

A Inteligência do Negócio (*Business Intelligence* – BI) é usada para compreender as capacidades disponíveis na empresa, o estado da arte, tendências e direções futuras nos mercados, as tecnologias e o ambiente regulatório no qual a empresa compete, bem como as ações dos concorrentes e as implicações dessas ações. O objetivo é de melhorar a pontualidade e qualidade dos insumos para o processo de decisão (Negash, 2004).

O surgimento de uma nova forma de competição baseada no uso intensivo de análise, dados e tomada de decisões baseada em fatos fez com que, no lugar de competir em fatores tradicionais, as empresas começassem a empregar estatística, análise quantitativa e modelagem preditiva como elementos primários de concorrência (Davenport et al, 2006).

Uma nova ciência, denominada Ciência de Dados (*Data Science*), começa a ganhar corpo, cuja definição continua a ser desenvolvida ao longo dos últimos anos, mas que, em suma, trata justamente desta combinação de habilidades e áreas de conhecimento, visando a coleta, preparação, análise, visualização, gerenciamento e preservação de grandes quantidades de informação.

O objetivo deste artigo é apresentar os diversos conceitos e técnicas de Inteligência de Negócio a Ciência de Dados, bem como suas potenciais aplicações, traçando um paralelo entre esses dois termos contemporâneos.

Este artigo está organizado da seguinte forma: a Seção 2 descreve conceitos relevantes de BI como uma classe de sistemas de suporte a decisão; a Seção 3 aborda a Ciência de Dados, a partir de alguns de seus conceitos, além de processos, tarefas e técnicas; a Seção 4 faz um estudo comparativo entre Ciência de Dados e outras áreas relacionadas, cujos objetivos são a tomada de decisão e a descoberta de conhecimento a partir de dados; a Seção 5 apresenta o potencial de aplicações da Ciência de Dados através de exemplos; e a Seção 6 contém a conclusão do artigo.

## **2. Business Intelligence e Sistemas de Suporte à Decisão**

BI – *Business Intelligence* é considerada como o processo de transformar dados em informação e então em conhecimento, permitindo a tomadores de decisão refinar suas ações de acordo com a estratégia do negócio (Golfarelli et al, 2004). BI é também uma das técnicas para melhorar, qualitativa e quantitativa, o valor do conhecimento disponível para tomadores de decisão (Cody et al, 2002). Os sistemas de BI combinam dados operacionais com ferramentas analíticas para apresentar informações complexas e competitivas para os planejadores e tomadores de decisão (Negash, 2004).

Barbieri (2011) apresenta uma definição mais prática de BI, tendo como foco sua aplicação pelas empresas, ao afirmar que o conceito de BI, de forma mais ampla, pode ser entendido como a utilização de variadas fontes de informação para definir estratégias de competitividade nos negócios da empresa.

O processo de BI envolve a aplicação de métodos de pesquisa, modelos e abordagens para vários problemas de domínios. O campo da gestão tem uma necessidade crescente de análise de dados (Moutinho e Huarng, 2012).

Historicamente, a primeira menção a BI foi feita em 1865, por Richard Millar Devens ([https://en.wikipedia.org/wiki/Business\\_intelligence](https://en.wikipedia.org/wiki/Business_intelligence)). Já nos tempos modernos, ao final dos anos 50, o pesquisador da IBM, Hans Peter Luhn, definiu BI como a “capacidade de apreender as inter-relações dos fatos apresentados de tal forma a orientar a ação em direção a um objetivo desejado” (Luhn, 1958). Em 1989, Howard Dresner (posteriormente analista do Gartner Group e considerado como “pai do BI”) propôs BI como um ‘termo guarda-chuva’, ou seja, uma ampla categoria de software e soluções para reunir, consolidar, analisar e fornecer acesso a dados de uma forma a permitir que os usuários corporativos possam tomar melhores decisões nos negócios.

Ao se abordar BI, ainda no campo conceitual e histórico, é fundamental mencionar os Sistemas de Suporte à Decisão (do Inglês *Decision Support Systems - DSS*), sendo estes considerados como precursores do primeiro, tendo sua definição e estudo sendo realizados ao longo dos últimos 40 anos.

A fim de melhor entender a relação entre BI e DSS, pode-se dizer que: DSS são sistemas ou subsistemas interativos, baseados em computadores, direcionados a auxiliar a tomadores de decisão a usar tecnologias de comunicação, dados, documentos, conhecimento e/ou modelos para identificar e resolver problemas, realizar tarefas completas do processo de decisão e a tomar decisões. DSS é um termo geral para qualquer aplicação de computador que melhora a habilidade de tomar decisões. Adicionalmente, DSS se refere a um campo acadêmico de pesquisa que envolve a concepção e o seu estudo no contexto de sua utilização. Os cinco tipos mais específicos de Sistemas de Suporte à Decisão incluem: DSS dirigido à comunicação; dirigido a dados; dirigido à documentação; dirigido ao conhecimento; e dirigido a modelos (Power, 2002).

Com base nos tipos de DSS, Power traça o paralelo entre BI e DSS, ao concluir que “BI é um DSS orientado a dados que apoia, principalmente, a consulta de uma base de dados histórica e a elaboração de relatórios periódicos resumidos – podendo, ainda, ser referido a um propósito específico para alguns DSS orientados a Dados”.

Atualmente, o termo BI está associado também a *Analytics*, privilegiando a capacidade de análise de dados das plataformas de BI, formando a nova sigla BI&A, de *Business Intelligence and Analytics* (Gartner, 2015).

### 3. Ciência de Dados

A Ciência de Dados (*Data Science*) é um conjunto de princípios fundamentais que suportam e guiam a extração de informações e conhecimento a partir de dados. Neste sentido, a Ciência de Dados envolve princípios, processos e técnicas para compreender fenômenos através da análise automatizada de dados (Provost e

Fawcett, 2013a).

Se *Ciência* é um método sistemático onde pessoas estudam e explicam fenômenos de um domínio específico que ocorrem no mundo natural, pode-se dizer que a Ciência de Dados é o domínio científico que é dedicado para descobrir conhecimento (*knowledge discovery*) através da análise de dados.

Os termos atuais, mais comuns, usados para definir a Ciência de Dados têm sido utilizados, ao longo das últimas décadas, ligados a outras áreas de conhecimento, como Análise de Dados, Processamento de Dados, Estatística, Descoberta de Conhecimento em Banco de Dados (KDD), Mineração de Dados (Data Mining), *Big Data*, entre outros. Logo, é possível perceber se tratar de um campo interdisciplinar, que divide definições e áreas de atuação com outros campos.

Uma ilustração da interdisciplinaridade envolvida na Ciência de Dados pode ser vista no “Diagrama Venn de Ciência de Dados” (Conway, 2010), ilustrado na Figura 1.

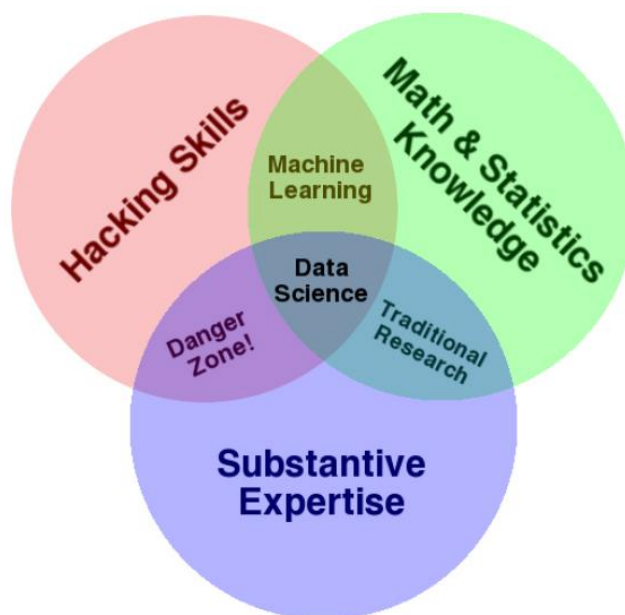


Figura 1. Diagrama Venn de Ciência de Dados (Conway, 2010)

Segundo Conway (2010), a aplicação, em conjunto, das habilidades descritas no Diagrama Venn – habilidades computacionais, conhecimento estatístico e matemático, e expertise no domínio/assunto – definem, na prática, a Ciência de Dados.

A fim de introduzir o assunto a ser tratado na próxima seção deste artigo, e fazendo referência ao que foi exposto acima, cabe citar o que diz Dhar (2013), ao enumerar alguns conceitos chaves relacionados à ciência de dados:

- Ciência de Dados é o estudo da extração de conhecimento generalizável a

partir de dados.

- Um requerimento epistêmico comum, para avaliar se um novo conhecimento é susceptível como recurso para a tomada de decisão é o seu poder preditivo e não apenas sua capacidade de explicar o passado.
- Um cientista de dados requer um conjunto de habilidades integradas que abrange matemática, aprendizagem de máquina, inteligência artificial, estatística, bases de dados e otimização, juntamente com uma profunda compreensão do ofício de formulação de um problema para projetar soluções eficazes.

#### **4. Ciência de Dados e outros campos – um estudo comparativo entre conceitos relacionados à tomada de decisão orientada a dados**

Atualmente, empresas das mais diversas indústrias, e também governos, se veem em meio a uma vasta quantidade de dados, disponível em diversas formas. Com isto, ambos focam suas atenções para a exploração de tais dados com o objetivo de atingir vantagens competitivas. Esta grande quantidade e variedade de dados trazem consigo novos desafios, uma vez que não é possível manuseá-los e analisá-los de forma manual. Por outro lado, vindo ao encontro desta mudança, os computadores e outros recursos relacionados, como capacidade de transmissão, armazenamento de dados e etc., tornam-se cada vez mais potentes. Segundo Provost e Fawcett (2013b), a convergência desse fenômeno deu origem a cada vez mais generalizada aplicação, nos negócios, da Ciência de Dados.

Como dito anteriormente, o conceito de Ciência de Dados vem sendo elaborado ao longo dos últimos anos. Neste sentido, há, inclusive, programas e discussões acadêmicas sendo desenvolvidos, em universidades e instituições de pesquisa. Por outro lado, ao ligarmos este tema ao mundo dos negócios, muito mais importante do que sua definição, é entender sua relação com outros conceitos importantes e intimamente relacionados. Nas próximas subseções, serão discutidos alguns desses conceitos, com foco em posicioná-los em relação à Ciência de Dados.

##### **4.1. Ciência de Dados e Inteligência de Negócios e Análise (BI&A)**

A Ciência de Dados (*Data Science*) e a Inteligência de Negócios (*Business Intelligence*) têm em comum suas principais funções: converter dados brutos em *insights* a serem usados no processo de tomada de decisões nos negócios ou domínios de aplicação em geral. Por outro lado, diferem em termos de suas abordagens.

O termo BI tornou-se popular nas comunidades de negócios e tecnologia da informação a partir dos anos 90. Na década seguinte, foi introduzida a análise de negócios (*Business Analytics*), a fim de representar o componente analítico chave em BI (Davenport et al, 2006). As técnicas analíticas de BI, normalmente aplicadas pelas organizações sobre sistemas legados, cujos dados são armazenados em bancos de

dados relacionais, são fundamentadas, principalmente, em métodos estatísticos, além de também utilizarem técnicas de mineração de dados. Com isso, e em adição aos já tradicionais relatórios típicos de BI, a análise estatística, combinada com a mineração de dados, é adotada para a análise de associações, segmentações e agrupamentos de dados, classificação e análise de regressão, detecção de anomalias e modelagem preditiva em várias aplicações comerciais (Chen et al, 2012). Neste sentido, empregam-se a Análise Descritiva, que descreve quantitativamente as características principais de uma coleção de dados, e a Análise Preditiva, que se concentra na análise correlativa, prevendo relações entre variáveis aleatórias conhecidas ou conjuntos de dados, a fim de identificar como um evento ocorrerá no futuro.

Em BI, embora métodos possam ser usados para ‘previsão de futuro’, os mesmos são gerados a partir de inferências simples tendo como base dados históricos ou atuais, a fim de oferecer informações relevantes, a curto ou médio prazos, aos tomadores de decisão.

Na Ciência de Dados, em contraste, procura-se fazer novas descobertas utilizando-se, para tal, métodos matemáticos ou estatísticos avançados para analisar e gerar previsões a partir de grandes quantidades de dados do negócio. Estes *insights* preditivos são geralmente relevantes para o futuro do negócio, a longo prazo.

Em linhas gerais, há significantes diferenças entre *BI&A* e *Data Science*. Cabe ressaltar que, embora diferentes, um não substitui o outro, pois, na verdade, são complementares.

A Tabela 1, a seguir, ressalta algumas das diferenças mencionadas acima.

Área	BI	Data Science
Tecnologia	Opera sobre bancos de dados relacionais, <i>data warehouses</i> , OLAP e tecnologias ETL.	Opera sobre dados provenientes de sistemas de engenharia de dados que usam Hadoop, MapReduce ou Processamento Paralelo.
Expertise	Depende fortemente de TI e de expertise em tecnologia de negócios.	Baseia-se em experiência em estatística, matemática, programação e negócios.
Foco	Relatórios, KPIs, tendências	Padrões, correlações, modelos
Processos	Estático, comparativo	Exploratório, experimentação, visual
Fontes de Dados	Planejada, adicionada	<i>On the fly</i> , conforme necessário

	lentamente	
Transformação	Prévia, cuidadosamente planejada	No banco de dados, conforme demanda, enriquecimento
Qualidade de Dados	Única versão da verdade	"Good enough", probabilidades
Modelo de Dados	Esquema de carga	Esquema em consulta
Análise	Retrospectiva, descritiva	Preditiva, prescritiva

Tabela 1: Diferenças entre BI e Data Science – Adaptado de (Schmarzo, 2014)

#### 4.2. Ciência de Dados e Engenharia de Dados

Ao relacionar estas áreas, temos, de um lado, a Ciência de Dados como a prática do uso de métodos computacionais para derivar informações valiosas a partir de conjuntos de dados brutos. Do outro lado, a Engenharia de Dados (*Data Engineering*) é um domínio da Engenharia dedicado a superar gargalos de processamento e problemas de manipulação de dados para aplicações que utilizam grandes volumes, variedades, e velocidades de dados. Enquanto a Ciência de Dados tem foco nos métodos computacionais, na captura e coleção dos dados em si, a Engenharia de Dados está focada em também manipular tais dados, mas direcionada à infraestrutura, modelagem, projeto e visualização das bases de dados.

Em ambas as áreas, é comum trabalhar com as três variações de dados a seguir:

- Dados estruturados: dados armazenados, processados e manipulados em um sistema de gerenciamento de banco de dados (SGBD) tradicional;
- Dados não estruturados: dados normalmente gerados por atividades manuais e que não se encaixam em um formato de banco de dados estruturado (e-mails, documentos textuais diversos);
- Dados semiestruturados: dados que não se encaixam em um formato de banco de dados estruturado, mas que possuem uma estruturação por *tags* que são úteis para a criação de uma forma de ordenação e hierarquia nos mesmos (documentos em formato XML).

Por fim, há ainda outro importante paradigma em comum nestas duas áreas – o paradigma de *Big Data* – uma abordagem em que enormes velocidades, variedades

e volumes de dados estruturados, não estruturados e semiestruturados são capturados, processados, armazenados e analisados.

Como dito acima, *Big Data* é um conceito relacionado a dados que excedem a capacidade de sistemas tradicionais de bancos de dados, devido às suas principais particularidades, como possuírem enormes volumes, se moverem rapidamente e não se encaixarem nos requisitos tradicionais de arquiteturas de bancos de dados. Quatro principais características – os Quatro V’s – definem o que é Big Data: volume, velocidade, variedade e valor.

Com o crescimento contínuo dos “quatro V’s”, soluções e tecnologias inovadoras de dados precisam ser continuamente desenvolvidas. Desde soluções mais simples, embora caras, como a aquisição de computadores com grande quantidade de memória e números de processadores – e que podem ser escalados pela adição, conforme a necessidade, de novas máquinas – onde o limite é o orçamento disponível, passando por soluções alternativas como clusters de alta disponibilidade – que têm em comum com a primeira alternativa seu alto custo, chegando, por fim, a soluções computacionais mais econômicas, como a computação em nuvem (*Cloud Computing*). Nesta última solução, um padrão comum é ter uma grande quantidade de dados para serem transformados, onde o processamento de cada item de dados é essencialmente independente dos demais itens. Nesse caso, um algoritmo de instrução única de dados múltiplos (SIMD) é utilizado.

Entre os algoritmos disponíveis para Big Data, cabe destacar o Hadoop – um framework de código aberto para computação em nuvem, que conta ainda com um sistema de arquivos distribuídos. Como uma de suas principais características, tal algoritmo suporta o modelo MapReduce, que foi introduzido pelo Google como um método para resolver uma classe *petascale* (operações de ponto flutuante superiores a um quatrilhão por segundo) de problemas com grandes *clusters* de máquinas de baixo custo (Venner, 2009).

Este modelo baseia-se em duas etapas distintas para uma aplicação: *Map*: uma etapa inicial de ingestão e transformação, na qual os registros individuais de entrada podem ser processados em paralelo; *Reduce*: uma etapa de agregação ou resumo, em que todos os registros associados devem ser processados juntos por uma única entidade.

Por fim, é possível ver o termo *Big Data* ser continuamente relacionado à Ciência de Dados (Dhar, 2013). Entretanto, cabe ressaltar que o termo possui papel fundamental em outras áreas relacionadas ao gerenciamento de dados, como a engenharia de dados e a mineração de dados.

#### **4.3. Ciência de Dados e Descoberta de Conhecimento em Bases de Dados**

Em um nível abstrato, a área de Descoberta de Conhecimento em Bases de Dados (do





Inglês, *Knowledge Discovery in Databases* – KDD) está voltada para o desenvolvimento de métodos e técnicas para dar sentido a dados. O problema básico endereçado pelo processo de KDD é o mapeamento de dados de baixo nível (que são tipicamente muito volumosos para se entender e sintetizar facilmente) em outras formas que possam ser mais compactas (por exemplo, um breve relatório), mais abstratas (por exemplo, uma aproximação descritiva ou modelo do processo que gerou os dados) ou mais úteis (por exemplo, um modelo preditivo para estimar o valor de casos futuros). No cerne desse processo está a aplicação de métodos específicos de mineração de dados para descoberta e extração de padrões (Fayyad et al, 1996). A Mineração de Dados (*Data Mining*) pode ser definida como o processo, com fases definidas, de extração informação a partir de conjuntos de dados e sua transformação em uma estrutura compreensível para uso futuro. Embora comumente relacionada à KDD, tal processo pode ser usado isoladamente e também em conjunto com outros métodos ou técnicas.

Além deste processo principal, há ainda outros cinco que compõem o KDD:

- Preparação de dados;
- Seleção de dados;
- Limpeza de dados;
- Incorporação do conhecimento prévio adequado;
- Interpretação correta dos resultados de mineração.

Em termos de métodos e técnicas, como se verá na conclusão deste artigo, a área de KDD se confunde com o que se tem definido como Ciência de Dados.

## 5. Aplicações da Ciência de Dados

Enquanto ciência pura, a Ciência de Dados se preocupa em aprimorar técnicas, modelos e métodos computacionais, matemáticos e estatísticos que possam influenciar nas ferramentas para a descoberta das demais áreas do conhecimento nas quais pretende ser aplicada. Para a obtenção de resultados concretos ao longo desta interação com as áreas de aplicação, parte-se da premissa de que é preciso detectar de forma precisa qual o problema subjacente que se pretende resolver. São as perguntas do pesquisador ou do gerente de negócios (questões de pesquisa) que constituem a bússola que impulsiona o desenvolvimento de novas teorias e ferramentas para a extração do conhecimento.

Dentro desse contexto, a Ciência de Dados vem oferecendo suporte a diversas áreas, tais como: finanças, medicina, astronomia, jogos, marketing, biologia, dentre inúmeras outras que dela vem se apoiando para responder a problemas científicos ou problemas relacionados à detecção de padrões de comportamento, como aqueles

voltados para a área de negócios. Neste sentido, a Ciência de Dados não deixa de ser uma extensão a BI quando aplicada a áreas de negócios.

Nos ramos das Ciências Naturais, uma das primeiras aplicações foi na Astronomia, alcançadas pelo sistema SKICAT [Fayyad, 1995].

Avanços extraordinários têm-se obtido no campo da genômica, desde o sequenciamento do genoma humano e de outras espécies de animais e plantas, cujas informações genéticas encontram-se depositadas em bancos de dados públicos, aguardando processamento para ganharem sentido. O crescente interesse por análises genômicas *in silico*, utilizando sistemas KDD, é resultado do sucesso que vem sendo alcançado por meio de tais análises na identificação de genes e suas funções, cuja sequência de nucleotídeos está disponível em bases de dados públicas na web.

Métodos estatísticos, capazes de analisar grandes quantidades de dados biológicos, vêm sendo desenvolvidos e implementados há anos na tentativa de identificar e prever as funções dos genes e proteínas por eles codificados (Wang, 2003). Resultados alcançados por meio de análise *in silico* e confirmadas posteriormente em laboratório (Silva, 2010), sugerem que é possível a identificação de uma família inteira de genes, por meio de processos de Ciência de Dados, a partir de informações genéticas depositadas em banco de dados públicos.

Recentes aplicações de Ciência de Dados, seja como extensão a BI ou como ciência aplicada, incluem domínios como Dados Abertos Governamentais (Veiga e Guimarães, 2015) e Hidrologia (Xavier et al, 2015), dando mostras da gama diversificada de áreas de conhecimento em que a Ciência de Dados pode ser aplicada.

## 6. Conclusão

Historicamente, a noção de descobrir padrões úteis em dados tem ganhado uma variedade de nomes, como: Mineração de Dados, Extração de Conhecimento, Descoberta de Informação, Coleta de Informação, Arqueologia de Dados e Processamento de Padrões de Dados. O termo Mineração de Dados tem sido mais usado por estatísticos, analistas de dados e por comunidades de gestores de sistemas de informação. O termo “descobrimto de conhecimento em bases de dados” foi cunhado no primeiro Workshop de KDD, realizado em 1989, para enfatizar que o conhecimento é o produto final da descoberta orientada a dados.

Tendo sido apresentados os conceitos relacionados a vários campos, e cumprindo com o objetivo deste artigo, conclui-se que a Ciência de Dados é uma área atual, interdisciplinar, recém-formada, cuja definição e conceitos continuam a ser elaborados, e que possui similaridades, assim como compartilha processos e técnicas, com várias outras áreas e processos já estabelecidos. Dentro destas similaridades, merecem destaque as semelhanças de Ciência dos Dados com a área de KDD, a ponto de não se identificar, no presente momento, características marcantes que



diferenciem definitivamente as duas áreas, quer seja nos métodos e técnicas utilizadas ou nos domínios de suas aplicações.

### Referências

- Barbieri, C. (2011) BI2 – Business Intelligence: Modelagem e Qualidade. Editora Elsevier.
- Chen, H., Chiang, R. H. L., Storey, V. C (2012), Business Intelligence and Analytics: from big data to big impact. In: MIS Quarterly, Vol. 36 No. 4, pp. 1165-1188.
- Cody, W. F., Kreulen, J. T., Krishna, V., Spangler, W.S. (2002), “The integration of business intelligence and knowledge management” IBM Systems Journal, Vol. 41/4.
- Conway, D. (2010), “The Data Science Venn Diagram”. Disponível em <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram> . Último acesso em maio de 2015.
- Davenport, T. H., Cohen, D., Jacobson, A. (2006), “Competing on Analytics” In: Harvard Business Review, p. 98-107.
- Dhar, V. (2013) “Data Science and Prediction” In: Communications of the ACM, Vol. 56 No. 12, pp. 64-73.
- Fayyad, U., Piatetsky-Shapiro, G., Smyth P., “From Data Mining to Knowledge Discovery in Databases”, 1996, In: Advances in Knowledge Discovery and Data Mining , pp. 1–34.
- Fayyad, Usama M. SKICAT: sky image cataloging and analysis tool. Proceedings of the 14th international joint conference on Artificial intelligence, 1995
- Gartner (2015). Magic Quadrant for Business Intelligence and Analytics Platforms. Disponível em <https://www.gartner.com/doc/2989518/magic-quadrant-business-intelligence-analytics>. Último acesso em agosto de 2015.
- Godin, Benoît. The Knowledge-Based Economy: Conceptual Framework or Buzzword? Journal of Technology Transfer, v. 31, no 1, p. 17-30, Jan. 2006.
- Golfarelli, M. and Rizzi, S. and Cella, I. (2004), “Beyond data warehousing: what’s next in business intelligence?” In: ACM International Workshop on Data Warehousing and OLAP, pp. 001-006.
- Luhn, H.P. (1958). "A Business Intelligence System" (PDF). IBM Journal 2 (4)
- Moutinho, L. and Huarng, K.-H. (Eds) (2012), “Quantitative Modelling in Marketing and Management” In: World Scientific, Singapore.
- Negash, Solomon. Business Intelligence. Communications of the Association for Information Systems (Volume13, 2004) 177-195
- Power, Daniel. J., Decision Support Systems: Concepts and Resources for Managers, Quorum Books, 2002.
- Provost, F., Fawcett, T. (2013a), “Data Science and its relationship to big data and data-driven decision making” In: Big Data Journal, Vol. 1, pp. 51-59.
- Provost, F., Fawcett, T. (2013b), “Data Science for Business”, O’Reilly Media.
- Schmarzo, B., “Business Intelligence Analyst or Data Scientist? What’s the Difference?”, 2014, Disponível em: <https://reflectionsblog.emc.com/2014/08/business-intelligence-analyst-data->



[scientist-whats-difference](#). Último acesso em junho de 2015.

- Silva, V.A. Determinação da estrutura organizacional das vias MAP KINASES em sorgo, *Arabidopsis lyrata* e cana-de-açúcar por meio de análise de Bioinformática. Tese de Doutorado. UENF Darcy Ribeiro, Agosto 2010.
- Veiga, J., Guimarães, J. Análise de Dados Abertos Governamentais usando Técnicas de Business Intelligence: um Estudo de Caso das Eleições 2014. Trabalho de Conclusão de Curso de Bacharelado em Sistemas de Informação, UNIRIO, 2015.
- Venner, J. (2009), "Pro Hadoop", New York, NY: Apress.
- Wang, D., Jeffrey F. Harper, and Michael Gribskov (2003) - Systematic Trans-Genomic Comparison of Protein Kinases between *Arabidopsis* and *Saccharomyces cerevisiae*-*Plant Physiology*, Vol. 132, August 2003.
- Xavier, F., Tanaka, A. K., and Revoredo, K. C. (2015). Application of knowledge discovery in databases in evapotranspiration estimation: an experiment in the State of Rio de Janeiro. In SBSI 2015 Proceedings, <http://aisel.aisnet.org/sbis2015/76>.