



**UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ**  
**DIRETORIA DE PESQUISA E PÓS-GRADUAÇÃO**  
**ESPECIALIZAÇÃO EM ENGENHARIA DE PRODUÇÃO**



**ATILAS FERREIRA DE PAIVA**

**APLICAÇÃO DE DATA SCIENCE COMO FERRAMENTA DE  
APOIO A TOMADA DE DECISÃO ORIENTADA POR DADOS**

**MONOGRAFIA DE ESPECIALIZAÇÃO**

**PATO BRANCO**

**2018**

**ATILAS FERREIRA DE PAIVA**

**APLICAÇÃO DE DATA SCIENCE COMO FERRAMENTA DE  
APOIO A TOMADA DE DECISÃO ORIENTADA POR DADOS**

Monografia apresentada como requisito parcial à obtenção do título de Especialista na Pós-Graduação em Engenharia de Produção da Universidade Tecnológica Federal do Paraná – UTFPR – *Campus* Pato Branco.

Orientador(a): Prof. Dr. Gilson Adamczuk Oliveira

**PATO BRANCO**

**2018**



## **TERMO DE APROVAÇÃO**

### **APLICAÇÃO DE *DATA SCIENCE* COMO FERRAMENTA DE APOIO A TOMADA DE DECISÃO ORIENTADA POR DADOS**

Por

**ATILAS FERREIRA DE PAIVA**

Este Trabalho de Conclusão de Curso (TCCE) foi apresentado em 23 de novembro de 2018 como requisito parcial para a obtenção do título de Especialista em Engenharia de Produção. O candidato foi arguido pela Banca Examinadora composta pelos professores abaixo assinados. Após deliberação, a Banca Examinadora considerou o trabalho aprovado.

---

Prof. Dr. Gilson Adamczuk Oliveira  
Prof.(a) Orientador(a)

---

Prof. Dr. José Donizetti de Lima  
Membro titular

---

Prof. Dr. Sergio Luiz Ribas Pessa  
Membro titular

- O Termo de Aprovação assinado encontra-se na Coordenação do Curso -

Dedico este trabalho à Deus quem fortalece, à minha esposa e meu filho que fazem dos meus dias melhores.

## **AGRADECIMENTOS**

Primeiramente, agradeço à Deus pela vida e a oportunidade de concluir mais uma etapa acadêmica, à minha esposa que sempre me incentivou estando ao meu lado me apoiando acreditando em mim.

Agradeço ao meu filho, o pequeno Samuel que chegou durante o período da especialização, trazendo alegria e motivação em prosseguir, tornando meus dias melhores.

Agradeço ao meu orientador Prof. Dr. Gilson Adamczuk Oliveira, quem prontamente por meio de uma ligação me direcionou a buscar cursar a especialização, a qual me despertou a buscar além daquilo que eu já julgava ser suficiente em conhecimento.

Gostaria de agradecer a todos os professores e colegas de turma, pela grande e valiosa troca de conhecimento e experiências.

## RESUMO

PAIVA, Atilas Ferreira. **APLICAÇÃO DE DATA SCIENCE COMO FERRAMENTA DE APOIO A TOMADA DE DECISÃO ORIENTADA POR DADOS.** 2018. 24 f. Monografia (Especialização em Engenharia de Produção) - Universidade Tecnológica Federal do Paraná. Pato Branco, 2018.

Atualmente várias empresas tem investido cada vez mais recursos e esforços para se tornar competitivas, onde muitas das vezes o “se tornar competitivo” e o “se manter competitivo” não é uma tarefa fácil a ser realizada, devido aos desafios econômicos e às exigências impostas pelo mercado consumidor, por produtos e serviços com cada vez mais qualidade, diversidade, disponibilidade e baixo custo. Almejando alcançar esta competitividade, crescimento e a sobrevivência de suas empresas, gestores estão buscando otimizar seu modelo de gestão, por meio de ferramentas e recursos que proporcionem informações cada vez mais precisas, que os apoiem e auxiliem na tomada de decisão, buscando a assertividade em seus investimentos e encontrar a melhor direção para a elaboração de suas estratégias. Como resposta a esta necessidade, surgiu então a Ciência de Dados ou *Data Science*, ciência esta que busca estudar, analisar e compreender grandes volumes de dados, por meio de estatística e tecnologias como *Data Warehouse*, *Big Data*, *Data Analytics*, *Machine Learn*, *Data Mining*, linguagens de programação e o conhecimento de regras de negócio. Em paralelo ao cenário de alta competitividade, observa-se a existência de um fenômeno de ascensão dos dados, caracterizado pelo aumento do volume de dados de forma exponencial, proporcionando uma ótima oportunidade para se adotar ao método de tomada de decisão orientada por dados, no qual o conhecimento e o domínio de determinadas tecnologias de mineração e análise de dados, somado ao conhecimento das regras de negócio, apoiarão aos gestores durante os processos decisórios, fazendo com que *Data Science* seja uma grande aliada no processo de implementação deste método. Este trabalho tem como objetivo apresentar os conceitos de *Data Science* bem como as principais tecnologias envolvidas e sua aplicação no que se refere ao processo de decisão orientada por dados.

**Palavras-chave:** *Data Science*, Tomada de decisão, Dados, Análise de dados, *Insight*, Informação, Conhecimento.

## ABSTRACT

PAIVA, Atilas Ferreira. **APPLICATION OF DATA SCIENCE AS SUPPORT TOOL THE DATA-ORIENTED DECISION MAKING..** 2018. 24 f. Monografia (Especialização em Engenharia de Produção) - Federal Technology University - Paraná. Pato Branco, 2018.

Today, a number of companies have invested more and more resources and efforts to become competitive, where often "becoming competitive" and "staying competitive" is not an easy task to be undertaken due to economic challenges and imposed requirements by the consumer market, by products and services with ever more quality, diversity, availability and low cost. Aiming to achieve this competitiveness, growth and survival of its companies, managers are seeking to optimize their management model, through tools and resources that provide increasingly accurate information, that support and assist in decision making, seeking assertiveness in your investments and find the best direction for the development of your strategies. As a response to this need, then came the Science of Data or Data Science, which seeks to study, analyze and understand large volumes of data, through statistics and technologies such as Data Warehouse, Big Data, Data Analytics, Machine Learn, Data Mining, programming languages and the knowledge of business rules. In parallel to the highly competitive scenario, there is a phenomenon of data ascension characterized by an exponential increase in data volume, providing an excellent opportunity to adopt the data-driven decision-making method in the the knowledge and mastery of certain mining technologies and data analysis, together with the knowledge of the business rules, will support the managers during the decision making process, making Data Science a great ally in the process of implementing this method. This paper aims to present the concepts of Data Science as well as the main technologies involved and their application with regard to the decision process driven by data.

**Keywords:** Data Science, Decision making, Data, Data analysis, Insight, Information, knowledge

## SUMÁRIO

|  |           |
|--|-----------|
| <b>1. INTRODUÇÃO .....</b>   | <b>02</b> |
| <b>2. DATA SCIENCE .....</b>   | <b>03</b> |
| 2.1. <i>O QUE É DATA SCIENCE? .....</i>  | 03        |
| <b>3. PRINCIPAIS TECNOLOGIAS APLICADAS EM DATA SCIENCE .....</b>                         | <b>04</b> |
| 3.1. <i>BANCO DE DADOS .....</i>   | 04        |
| 3.2. <i>DATA WAREHOUSE.....</i>  | 05        |
| 3.3. <i>BIG DATA.....</i>  | 06        |
| 3.4. <i>DATA MINING.....</i>   | 07        |
| 3.5. <i>MACHINE LEARNING.....</i>  | 08        |
| <b>4. TOMADA DE DECISÃO ORIENTADA POR DADOS .....</b>                                    | <b>09</b> |
| <b>5. APLICAÇÃO DE DATA SCIENCE NOS NEGÓCIOS.....</b>                                    | <b>10</b> |
| 5.1. <i>ESTUDO DE CASO DE APLICAÇÃO E DATA SCIENCE: WALMART E O FURACÃO FRANCES.....</i> | 12        |
| <b>CONCLUSÃO .....</b>   | <b>14</b> |
| <b>REFERÊNCIAS.....</b>  | <b>16</b> |



## 1. INTRODUÇÃO

Atualmente o mundo que está soterrado por dados. Os websites rastreiam todos os cliques de todos os usuários. Smartphones estão fazendo um registro de localização e sua velocidade a cada segundo diariamente. Atletas avaliados usam pedômetros com esteroides que estão sempre registrando suas batidas do coração, hábitos de movimentos, dietas e padrões do sono. Carros inteligentes coletam hábitos de direção, casas inteligentes hábitos de moradia e marqueteiros inteligentes coletam hábitos de compra. (GRUS, 2016). Segundo PROVOST e FAWCETT (2016), os últimos quinze anos testemunharam grandes investimentos em infraestrutura de negócios que têm melhorado a capacidade de coletar dados em toda a empresa.

No passado, as empresas podiam contratar equipes de estatísticos, modeladores e analistas para explorar manualmente os conjuntos de dados, mas seu volume e variedade superam muito a capacidade da análise manual. Ao mesmo tempo, computadores se tornaram muito mais poderosos, a comunicação em rede é onipresente, e foram desenvolvidos algoritmos que podem conectar conjunto de dados para permitir análise muito mais amplas e profundas do que antes. A convergência desses fenômenos deu origem à aplicação, cada vez mais difundida, de princípios de *Data Science* e de técnica de mineração de dados nos negócios. (PROVOST; FAWCETT, 2016).

Um ponto importante a ser destacado neste cenário de “Ascensão dos dados” mediante a competitividade do mercado, seria o processo de tomada de decisão por meio dos gestores, os quais se veem desafiados a tomar decisões cada vez mais precisas e assertivas, das quais muitas destas decisões, precisam ser tomadas em um curto espaço de tempo. Considerando o de cenários de alta competitividade ao qual as empresas estão inseridas, associado a necessidade dos gestores de serem precisos e assertivos em suas decisões, juntamente com o grande volume de dados e o surgimento de novas tecnologias que possibilitem a geração de informações, este trabalho tem como intuito, propor a utilização de *Data Science* ou Ciência de Dados como ferramenta de apoio e suporte à tomada de decisão. Sendo norteado este trabalho pelas seguintes perguntas: O que é *Data Science*? Quais as tecnologias envolvidas? Como utilizar *Data Science* como apoio à tomada de decisão?

## 2. DATA SCIENCE

### 2.1. O QUE É DATA SCIENCE?

De forma resumida, pode-se definir que *Data Science* é a ciência que busca extrair informação, conhecimento e *insight* a partir das mais variadas fontes volumes de dados, por meio da coleta, processamento e análise de dados, no qual este processo pode e poderá ser automatizados por meio de tecnologias da informação e métodos e modelos estatísticos e matemáticos.

Segundo Coelho (2017) *Data Science* ou Ciência de Dados é um estudo muito disciplinado com relação aos dados e demais informações inerentes à empresa e as visões que cercam um determinado assunto. Em resumo é uma ciência que visa estudar as informações, seu processo de captura, transformação, geração e, posteriormente, análise de dados. A ciência de dados envolve diversas disciplinas.

São elas:

- Estatística
- Computação
- Conhecimento do negócio
- Matemática

PROVOST e FAWCET (2016) afirmam que o objetivo primordial de *Data Science*, é o aprimoramento da tomada de decisão, uma vez que isso geralmente é interesse direto para os negócios.

Pode-se afirmar que *Data Science*, é a ciência que estuda os dados por meio de análises, buscando encontrar ou descobrir informações relevantes que possam vir a resolver problemas, questionamentos e gerar conhecimento.

Fazendo uma breve analogia entre *Data Science* e uma fábrica, assim como uma linha de produção necessita de matéria prima, mão-de-obra e equipamentos para se fabricar um determinado produto, com *Data Science* não é diferente, onde os dados são sua principal matéria-prima, as tecnologias e estatísticas bem como os profissionais destas áreas, podem ser comparados aos equipamentos e mão-de-obra que moldam e transformam os dados, visando chegar ao produto acabado que neste caso seria a informação ou *insight*.

Segundo PROVOST e FAWCET (2016) *Data Science* envolve princípios, processos e técnicas para compreender fenômenos por meio da análise (automatizada) de dados.

Ciência de dados é um campo interdisciplinar de investigação de dados que resolve problemas reais de negócios, com o uso de método científico e técnicas avançadas de análise de dados, *Machine Learning* e Inteligência Artificial. Esta é uma área essencial para posicionar as organizações no cerne da Indústria 4.0 (PRATES; HOPPEN, 2018)

### **3. PRINCIPAIS TECNOLOGIAS APLICADAS EM *DATA SCIENCE*.**

Conforme visto anteriormente, *Data Science* é uma ciência focada em extrair e gerar conhecimento por meio da análise e estudo sistemático dos dados, no entanto, considerando o crescimento exponencial do volume de dados em nossos dias atuais, a tecnologia da informação tem sido uma grande aliada desta ciência sendo um de seus principais pilares, possibilitando a automatização da coleta, armazenamento, tratamento e análise de grandes volumes de dados.

Dentre as tecnologias presente em *Data Science*, pode-se destacar como sendo as principais as seguintes: banco de dados, *Data Warehouse*, *Big Data*, *Data Mining*, *Machine Learning*, apresentados a seguir.

#### **3.1. BANCO DE DADOS**

Em uma definição mais ampla, um banco de dados é qualquer coisa que colete e organize dados. Uma planilha contendo reservas de clientes é um banco de dados, assim como um arquivo de texto simples contendo dados de horários de voos. Os próprios dados de teste simples podem ser armazenados em vários formatos, inclusive XML e CSV. Profissionalmente, no entanto, quando alguém se refere a um “banco de dados” provavelmente está se referindo a um *sistema de gerenciamento de banco de dados relacional* (RDBMS, *Relational Database Management System*) (NIELD, THOMAS, 2016).

Um Banco de Dados, pode ser definido como um local de armazenamento de dados, podendo este local ser físico, organizado e composto por pastas e arquivos impressos e locais lógicos ou virtuais sendo estes utilizados em *Data Science*, compostos Sistemas de Gerenciamento de Banco de Dados (SGBD) que utilizam um conjunto de tabelas para organizar dados eletrônicos de diversos tipos e formatos.

### 3.2. DATA WAREHOUSE

Segundo SINGH (2001, p. 12), *Data Warehouse* é o processo de integração dos dados corporativos de uma empresa em um único repositório a partir do qual os usuários finais podem facilmente executar consultas, gerar relatórios e fazer análises.

Um *Data Warehouse* é um ambiente de suporte à decisão que alavanca dados armazenados em diferentes fontes e os organiza e entrega aos tomadores de decisões da empresa, independentemente da plataforma que utilizam ou de seu nível de qualificação técnica. Resumidamente, *Data Warehouse* é uma tecnologia de armazenamento, gestão e análise de dados, coletados e pré-selecionados em outras fontes de dados.

O objetivo de um *Data Warehouse* segundo GONÇALVES (2013) é “armazenar os dados coletados de diferentes fontes (banco de dados operacionais, arquivos textos, sistemas legados, etc.) em um formato consistente.

Um *Data Warehouse*, semelhantemente ao Banco de Dados, também é um local de armazenamento de dados, no entanto, destina-se ao armazenamento de dados coletados e pré-selecionados em outras fontes de dados e aplicações de uma organização, afim de armazenar o que podemos chamar de “Supra Sumo” dos dados, com o objetivo de otimizar e agilizar a descoberta de conhecimento, informação e *insight*.

O Data Warehouse pode ser comparado ao Big Data, porém composto por um conjunto de dados restrito à organização a qual pertence, sendo um local de concentração de grandes volumes de dados coletados nas diversas fontes presentes dentro da organização.

### 3.3. BIG DATA

É um conceito que descreve o grande volume de dados estruturados e não estruturados que são gerados a cada segundo (NASCIMENTO, 2017).

Para a Oracle (2016) em um documento publicado em seu site *Big Data* é definida da seguinte forma: *Big data* descreve uma estratégia de gerenciamento de informações holística que incluem e integram muitos tipos de dados e gerenciamento de dados juntamente com dados tradicionais. Big data também foi definido pelos quatro V's:

**Volume:** A quantidade de dados. Enquanto volume indica *mais* dados, é a natureza granular dos dados que é única. Big data requer processamento de grandes volumes de dados *Hadoop* de baixa densidade e não estruturados – ou seja, dados de valor desconhecido, como *feeds* de dados do *Twitter*, fluxos de clicks em um site e em um aplicativo para dispositivos móveis, tráfego de rede, equipamento ativado por sensor que captura dados na velocidade da luz e muito mais. É tarefa de big data converter tais dados *Hadoop* em informações valiosas. Para algumas organizações, podem ser dezenas de terabytes, para outras podem ser centenas de petabytes.

**Velocidade:** A taxa rápida à qual os dados são recebidos e talvez administrados. A velocidade mais alta com que dados são transmitidos para a memória em relação à velocidade com que são gravados em disco. Alguns aplicativos de Internet das coisas (IoT – *Internet of Things*) possuem ramificações de saúde e segurança que requerem avaliação e ação em tempo real. Outros produtos inteligentes ativados por internet operam em tempo real ou quase em tempo real. Por exemplo, aplicativos de e-Commerce do consumidor buscam combinar preferências pessoais e de localização do dispositivo móvel para realizar ofertas de marketing sensíveis ao tempo. Operacionalmente, as experiências de aplicativos móveis possuem grandes populações de usuários, tráfego de rede aumentado e a expectativa de resposta imediata.

**Variedade:** Novos tipos de dados não estruturados. Tipos de dados não estruturados ou semiestruturados, como texto, áudio e vídeo, requerem

processamento adicional para derivar significado e metadados de apoio. Depois de interpretados, os dados não estruturados possuem muitos dos mesmos requisitos que os dados estruturados, como sumarização, linhagem, auditoria e privacidade. Mais complexidade surge quando os dados de uma fonte desconhecida mudam sem aviso. Mudanças do esquema frequentes ou em tempo real são um fardo enorme para ambientes de transação e análise.

**Valor:** Os dados possuem valor intrínseco – mas isso deve ser descoberto. Há uma série de técnicas quantitativas e investigativas para derivar valor a partir dos dados – desde a descoberta da preferência ou sentimento do cliente à realização de uma oferta relevante por localização ou para identificação de uma peça do equipamento que está prestes a falhar. A revolução tecnológica é que o custo do armazenamento e computação de dados diminuiu exponencialmente, oferecendo, assim, uma abundância de dados a partir de análises estatísticas em todo o conjunto de dados em relação a somente uma amostra anterior. A revolução tecnológica torna as decisões o mais precisas possível. Porém, encontrar valor também requer novos processos de descoberta que envolvem análises inteligentes e informativas, usuários comerciais e executivos. O verdadeiro desafio de big data é o humano, que está aprendendo a fazer as perguntas corretas, reconhecer padrões, fazer conjecturas informadas e prever comportamento.

De uma forma mais simples, pode-se definir *Big Data* como sendo o fenômeno atual de geração de dados não estruturados em larga escala.

### 3.4. DATA MINING

Para JEFFERSON (2011) *Data Mining* consiste em um processo analítico projetado para explorar grandes quantidades de dados (tipicamente relacionados a negócios, mercado ou pesquisas científicas), na busca de padrões consistentes e/ou relacionamentos sistemáticos entre variáveis e, então, validá-los aplicando os padrões detectados a novos subconjuntos de dados. O processo consiste basicamente em três etapas: exploração, construção de modelo ou definição do padrão e validação/verificação.

Segundo PROVOST e FAWCET (2016) A mineração de dados é uma arte. Como acontece com muitas artes, existe um processo bem definido que pode ajudar a aumentar a probabilidade de um resultado bem-sucedido. Esse processo é uma ferramenta conceitual fundamental para pensar sobre o projeto de *Data Science*.

*Data Mining* ou mineração de dados se resume explorar grandes volumes de dados, buscando descobrir padrões e extrair dados relevantes e específicos que possibilitem sua análise e conseqüentemente a geração de informações e conhecimentos.

Realizando um paralelo entre *Data Mining*, *Big Data* e *Data Warehouse*, constata-se que *Data Mining* tem por objetivo a exploração e a procura por dados relevantes em grandes volumes de dados, ou no próprio *Big Data*, armazenando estes dados em um *Data Warehouse*, caracterizado por ser um local de armazenamento de dados pré-selecionados coletados de outras fontes. Em outras palavras, *Big Data* é a matéria-prima à qual o *Data Mining* utiliza para extrair dados de maior relevância e os armazena no *Data Warehouse* para serem processados e analisados afim de gerar conhecimento e informação.

### 3.5. MACHINE LEARNING

*Machine Learning* ou Aprendizado de máquina, pode ser definida como a capacidade de aprendizado dos computadores, por meio da análise de dados históricos e a execução de instruções pré-programadas, possibilitando a tomada de decisão e ação por parte dos softwares e equipamentos.

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 (MATOS, 2015).

#### 4. TOMADA DE DECISÃO ORIENTADA POR DADOS

Tomada de Decisão Orientada por Dados (DOD) refere-se à prática de basear as decisões na análise de dados, em vez de apenas na intuição. Por exemplo, comerciante poderá selecionar anúncios baseados puramente em sua longa experiência na área e em sua intuição de que funcionará. Ou, pode basear sua escolha na análise dos dados sobre a forma como os consumidores reagem a diferentes anúncios. Ele também poderia utilizar uma combinação dessas abordagens. A DOD não é uma prática do tipo “tudo ou nada”, e diversas empresas a adaptam em maior ou menor grau (PROVOST; FAWCETT, 2016).

Os benefícios da tomada de decisão orientada por dados têm sido demonstrados conclusivamente. O economista Erik Brynjolfsson e seus colegas do MIT, e da *Penns Wharton School* realizaram um estudo de como DOD afeta o desempenho das empresas quanto ao uso de dados para tomar decisões. Eles mostram que, estatisticamente, quanto mais orientada por dados, mais produtiva uma empresa é – mesmo controlando o uma vasta gama de possíveis fatores de confusão. E as diferenças não são pequenas. Um desvio padrão a mais na escala de DOD está associado com um aumento de 4%-6% na produtividade. A DOD também está correlacionada com maior retorno sobre ativos, retorno sobre o patrimônio líquido, utilização de ativos e valor de mercado e a relação parece casual (PROVOST; FAWCETT, 2016).

Segundo HEINRICHSE (2013) para competir no mercado global de hoje, as empresas precisam deter mais conhecimento do que antigamente e, ainda, para obter sucesso, elas precisam saber mais sobre seus clientes, mercados, tecnologias e processos, e precisam ter essas informações antes que seus concorrentes.

FURLAN e FILHO, (2005) afirmam que ter informações em mãos é uma poderosa ferramenta para quem precisa tomar decisões e, tendo em vista esse princípio, as empresas começaram a extrair dados dos seus sistemas operacionais e armazená-los, separados dos dados operacionais.

A formulação estratégica de qualquer negócio sempre é feita a partir das informações disponíveis e, portanto, nenhuma estratégia consegue ser melhor que a informação da qual é derivada (REZENDE, 2002).

Para PATROCINIO (2017) atualmente existem dois cenários macro de aplicações de tomadas de decisão orientada por dados, usando os princípios



de *Data Science*: (1) as descobertas realizadas a partir de dados e (2) decisões repetitivas e em grande escala. O caso (1) está mais próximo do que hoje é chamado de *Advanced Analytics*, onde a empresa adquire novas informações apenas "olhando" para os dados. Grandes exemplos deste formato estão no Walmart, como o caso do furacão Frances e as associações de compra entre fraldas e cervejas e a mudança da disposição dos produtos em suas lojas físicas. Já o caso (2) pode ser representado pelos sistemas de recomendação, onde a própria aplicação decide automaticamente quais produtos deve exibir para o usuário.

## 5. APLICAÇÃO DE *DATA SCIENCE* NOS NEGÓCIOS

Analisando todo o contexto apresentado anteriormente, podemos afirmar que *Data Science* surgiu pela necessidade do mundo dos negócios por informações, pela busca dos gestores por conhecimento e *insight* que os apoiassem e dessem suporte no processo de tomada de decisão, seja para conquistar uma fatia maior do mercado, identificar oportunidades de redução de custos e aumento de lucros e que trouxessem alguma vantagem competitiva para suas empresas.

PROVOST e FAWCET (2016) afirmam que os dados e a capacidade de extrair conhecimento útil a partir deles, devem ser considerados importantes ativos estratégicos.

Conforme PRATES; HOPPEN (2018) Para aplicar projetos de ciência de dados em negócios, é fundamental cumprir algumas etapas, as quais são descritas a seguir.

### Definição de problema e métrica de sucesso

O ponto inicial para aplicar ciência de dados nas organizações é identificar qual é a dor de negócio, e qual é o indicador que realça essa dor na prática.

### Definição do *dataset* analítico a ser utilizado

O *dataset* (conjunto de dados) a ser utilizado pelo cientista de dados, não é uma simples extração de uma enorme massa de dados diretamente do banco transacional pelas queries SQL. (Saiba mais sobre projetos *analytics* de alto nível aqui.). O *dataset* analítico é construído a partir da definição do problema, e precisa conter todas as variáveis (colunas) necessárias para que o problema em questão

possa ser respondido. Este conjunto de dados normalmente é uma mescla de variadas bases de dados, tanto internas (da própria organização), quanto externas. Nas fontes de dados externas entram bases compradas de fornecedores específicos, ou até bases públicas, que têm o fim de enriquecer os dados e melhorar a assertividade dos modelos que serão aplicados.

### Transformação e higienização de dados

Este processo inclui pontos que envolvem técnicas estatísticas, como tratamento de dados faltantes e tratamento de outliers. Outros procedimentos também são realizados nesta etapa, como: “merge” ou concatenação de colunas, enriquecimento de dados com bases externas (dados de latitude e longitude, temperatura, macroeconômicos, entre outros) e diversas outras reestruturações necessárias para que os modelos de inteligência artificial consigam trazer a resposta desejada.

### Mineração de dados e modelagem com inteligência artificial

Um dos pontos mais importantes no processo de gerar valor para as companhias na área de ciência de dados é o processo de modelagem. Nesta etapa diversos modelos (centenas ou milhares) são treinados com o uso de diversas técnicas de inteligência artificial. O propósito de todas essas modelagens é encontrar as inter-relações entre as variáveis (colunas) com o uso de inteligência artificial, e gerar outputs como: predições: probabilidade de ocorrência de um evento dado um conjunto de características (leia mais sobre análise preditiva); previsões: projeções de como será o futuro de séries temporais; análise de perfis: identificação de personas, categorização entre indivíduos similares dentro do conjunto de dados, o que permite encontrar também grupos de outliers, os quais possuem características muito diferentes dos demais; criação de cenários: identificar grupos que impactam positiva ou negativamente em um target, uma variável a ser explicada.

### Comunicação dos resultados obtidos em linguagem de negócio

A comunicação dos resultados dos projetos de ciência de dados deve ser feita em linguagem de negócios, com foco na objetividade e na agilidade, mostrando os KPI's que foram impactados com o projeto e qual será o retorno financeiro obtido.

Muitas empresas consideram a análise de dados como pertencentes, principalmente, à obtenção de valores a partir de alguns dados existentes e, muitas vezes, sem preocupações com relação a se o negócio possui o talento analítico apropriado. Visualizar isso como ativos nos permite pensar explicitamente sobre a extensão em que se deve investir neles. Muitas vezes, não temos os dados corretos para melhor tomar decisões e/ou o talento certo para melhor apoiar a tomada de decisão a partir dos dados. (PROVOST; FAWCETT, 2016).

Analisando a citação a cima, vemos que a aplicação de *Data Science* como ferramenta de apoio à tomada de decisão dentro de uma organização, não depende somente do fato de se possuir dados para analisar, mas também se deve considerar como uma das principais “partes” recursos humanos com talentos destinados a exploração dos dados, ou seja, pessoas capazes de explorar os diversos volumes e fontes de dados com o objetivo de responder perguntas até então sem respostas, pessoas às quais dominem sim as tecnologias necessárias, mas que também entendam do negócio e saibam apresentar os resultados obtidos por meio do processo de análise dos dados.

A aplicação de *Data Science* pode ser utilizada para a realização de análises descritivas as quais buscam compreender situações e eventos em tempo real, análises preditivas que buscam prever cenários futuros por meio da identificação e análise de padrões de comportamento, análises prescritivas que podem indicar ações a serem tomadas mediante determinados cenários e situações, por fim a análise diagnóstica que busca compreender as causas de um comportamento ou evento e os detalhes real situação de determinados eventos.

## 5.1. ESTUDO DE CASO DE APLICAÇÃO E *DATA SCIENCE*: WALMART E O FURACÃO FRANCES

Considere um exemplo de uma história do *New York Times* de 2014

O FURACÃO FRANCES estava a caminho, avançando pelo Caribe, ameaçando atingir diretamente a costa atlântica da Flórida. Moradores de locais mais altos, mas longe, em Bentonville, Arkansas, executivos da

Wal-Mart Stores decidiram que a situação oferecia uma grande oportunidade para uma de suas mais novas armas baseadas em dados, algo que a empresa chama de tecnologia preditiva.

Uma semana antes da chegada da tempestade, Linda M. Dillman, diretora de informática do Wal-Mart, pressionou sua equipe a fazer previsões baseadas no que havia acontecido quando o furacão Charley atingiu várias semanas antes. Apoiada pelo trilhões de bytes de histórico de clientes armazenados na rede de computadores do Wal-Mart, ela sentiu que a empresa poderia "começar a prever o que vai acontecer, em vez de esperar que isso aconteça" (Hays, 2004).

Os especialistas extraíram os dados e descobriram que as lojas precisariam de certos produtos - e não apenas das lanternas comuns. "Não sabíamos no passado que os Pop-Tarts de morango aumentavam as vendas, como sete vezes a taxa normal de vendas, à frente de um furacão", disse Dillman em uma entrevista recente. "E o item mais vendido antes do furacão era cerveja".

Neste caso, pode-se observar que a aplicação de *Data Science* por meio de *Data Mining* e um conjunto de dados históricos, gerou informações que conduziu a empresa ao entendimento do comportamento dos clientes em uma determinada situação, possibilitando a previsão de comportamento futura mediante a situações iguais ou semelhantes, fazendo com que a empresa otimizasse as vendas de forma mais assertiva, impulsionando os produtos identificados por meio da análise dos dados com maior potencial histórico de vendas.

## 6. CONCLUSÃO

Em meio ao atual cenário de alta competitividade ao qual as empresas estão inseridas, seus gestores se veem desafiados à serem cada vez mais assertivos e precisos em suas decisões para elaborar estratégias com o objetivo de proporcionar crescimento e a sobrevivência de suas empresas, buscando alcançar uma maior lucratividade e redução de custos e ao mesmo tempo atendendo as exigências do mercado.

Ao mesmo tempo que vivemos em um contexto de alta competitividade entre as diversas empresas, vivemos também um tempo onde o crescimento do volume de dados tem ocorrido de forma exponencial, tornando este cenário uma grande oportunidade de extrair conhecimento, informações e *insights* por meio da análise destes dados utilizando *Data Science*. Sua utilização poderá possibilitar a realização de análises descritivas as quais buscam compreender situações e eventos em tempo real, análises preditivas que buscam prever cenários futuros por meio da identificação e análise de padrões de comportamento, análises prescritivas que podem indicar ações a serem tomadas mediante determinados cenários e situações e por fim a análise diagnóstica que busca compreender as causas de um comportamento ou evento e os detalhes real situação de determinados eventos.

A aplicação de *Data Science* ao processo de tomada de decisão, não deve ser entendida como uma automação deste processo em sua totalidade, mas sim como uma ferramenta de apoio e suporte, fazendo-se necessário o conhecimento das regras de negócio e do mercado a qual estão inseridos. *Data Science* não surgiu para eliminar a experiência, conhecimento e pensamento analítico por parte dos gestores, mas tem como objetivo ser uma aliada para gerar informações que tendem, de certa forma, facilitar e ajudar no processo de avaliação de cenários e possibilidades durante o processo de tomada de decisão.

A utilização de *Data Science* como ferramenta de apoio a tomada de decisão orientada à dados, traz consigo diversos benefícios, que vão desde identificar a saúde da empresa até mesmo identificar possíveis ações de que possibilitem um aumento das vendas, no entanto, sua aplicação e execução deve ser feita de forma planejada, com recursos humanos que componham uma equipe que possuam conhecimento e habilidades multidisciplinares, dominando as principais tecnologias

utilizadas em *Data Science* e conhecimento de das regras de negócio da empresa à qual está inserida.

## REFERÊNCIAS

PROVOST, FOSTER; FAWCETT, TOM. **Data Science para Negócios: O que Você Precisa Saber Sobre Mineração de Dados e Pensamento Analítico de Dados.** Rio de Janeiro: Alta Books, 2016.

GRUS, JOEL. **Data Science do Zero: PRIMEIRAS REGRAS COM O PYTHON.** Rio de Janeiro: Alta Books, 2016.

MATOS, DAVID. **O que é Data Science?**. Disponível em 2015 < <http://www.cienciaedados.com/o-que-e-data-science/> Acesso em 10 set. 2018.

COELHO, LUCAS. **CIÊNCIA DE DADOS: O QUE É, CONCEITO E DEFINIÇÃO.** Disponível em 2017 < <https://www.cetax.com.br/blog/data-science-ou-ciencia-de-dados/> > Acesso em 10 set. 2018.

NIELD, THOMAS. **Introdução à linguagem SQL: ABORDAGEM PRÁTICA PARA INICIANTEs.** São Paulo: Novatec, 2016.

FURLAN, MARCO R.; FILHO, EDUCARDO V. G. **UMA PROPOSTA DE APLICAÇÃO DE BUSINESS INTELLIGENCE NO CHÃO-DE-FABRICA,** GESTÃO & PRODUÇÃO, v.12, n.1, p.55-66, jan.-abr. 2005

HEINRICHs, J. H.; LIM, J. S. **Integrating web-based data mining tolls with business models for knowledge management.** *Decision Support Systems.* v. 35, n. 1, p. 103-112, 2003.

REZENDE, YARA. **Informação para negócios: os novos agentes do conhecimento e a gestão do capital intelectual.** *Ci. Inf.* [online]. 2002, vol.31, n.1, pp.75-83.

PATROCINIO, WESLEYS S. **Data Science é metodologia, e não tecnologia.** Disponível em 2017 < <https://medium.com/@wespatrocinio/data-science-%C3%A9-metodologia-e-n%C3%A3o-tecnologia-9c50fec73640> > Acesso em 15 set. 2018.

SINGH, HARRY S. **Data Warehouse: Conceitos, Tecnologias, Implementação e Gerenciamento.** 1ª ed. São Paulo: Makron Books, 2001.

GONÇALVES, MARCIO. **Extração de dados para Data Warehouse.** 1ª ed. Rio de Janeiro: Axcel Book, 2003.

JEFFERSON. **Conceitos e Técnicas sobre Data Mining.** Disponível em 2011 < <https://www.devmedia.com.br/conceitos-e-tecnicas-sobre-data-mining/19342> > Acesso em 13 set. 2018.

NASCIMENTO, RODRIGO. **Afinal, o que é Big Data?.** Disponível em 2017 < <http://marketingpordados.com/analise-de-dados/o-que-e-big-data-%F0%9F%A4%96/> > Acesso em 15 set. 2018.

**An Enterprise Architect's Guide to Big Data.** Oracle Disponível em 2016 < <http://www.oracle.com/technetwork/topics/entarch/articles/oea-big-data-guide-1522052.pdf> > Acesso em 15 set. 2018.

MATOS, DAVID. **Conceitos Fundamentais de Machine Learning.** Disponível em 2015 < <http://www.cienciaedados.com/conceitos-fundamentais-de-machine-learning/> > Acesso em 16 set. 2018.

PRATES, WLADEMIR; HOPPEN, JONI. **O que é ciência de dados (data science) e como aplicá-la nos negócios?.** Disponível em 2018 < <https://www.aquare.la/o-que-e-ciencia-de-dados-data-science-para-negocios/> > Acesso em 10 nov. 2018.

CONSTANCE L. HAYS. **What Wal-Mart Knows About Customers' Habits.** Disponível em 2004 < <https://www.nytimes.com/2004/11/14/business/yourmoney/what-walmart-knows-about-customers-habits.html> > Acesso em 22 nov. 2018.