

Nome do arquivo:

Tarefa01\_FilipeGutemberg\_74538\_JorgeMourão\_72799\_MatheusMarinho\_72753\_RodrigoGomes\_72594

## **TAREFA 01**

### **1. O que se entende por Ciência de Dados?**

Segundo [1], data science ou ciência de dados, é o estudo disciplinado, que se relaciona com dados e informações inerentes à empresa e as visões que cercam um determinado assunto.

Em resumo, é uma ciência que procura estudar as informações, passando pelas etapas de processo de captura, transformação, gestão e, posterior, análise de dados. A ciência de dados envolve diversas disciplinas, tais como: estatística, computação, conhecimento do negócio e Matemática [2].

Assim possibilita-se a extração de conhecimento, detecção de padrões ou obtenções de insights para possíveis tomadas de decisão.

### **2. O que faz um cientista de dados na indústria?**

Atualmente, grande parte dos negócios já fazem parte do mercado de dados e informações, sendo assim, está cada vez mais consistente no mercado para que possamos acompanhar este crescimento sem futuros inconvenientes.

Inevitavelmente, a maioria das empresas fazem parte desse ecossistema de dados, e sem um acompanhamento e análises cuidadosas num curto prazo de tempo, será quase impossível tomar decisões inteligentes a favor de bons resultados. É importante destacar que os dados permitem que as empresas e organizações entendam melhor os seus clientes e consumidores, e possam aprimorar cada vez mais os seus processos [2].

Sendo responsável por transformar dados em informações ou produtos de informações dentro de uma corporação, um cientista de dados também deve ser responsável pela formulação dos problemas, escolha de modelos de simulação e estatística e entrega dos produtos de dados.

Entre as suas principais atividades estão:

- Coletar grandes quantidades de dados não-estruturados e transforma-os num formato mais utilizável;
- Resolver problemas de negócios usando técnicas orientadas por dados;

- Trabalhar com uma variedade de linguagens de programação, incluindo SAS, R e Python;
- Ter uma sólida compreensão de estatística, incluindo testes e distribuições;
- Estar sempre atualizado sobre técnicas analíticas, como machine learning, deep learning e análise de texto;
- Comunicar e colaborar tanto com TI quanto com a gerência;
- Procurar por ordem e padrões nos dados, além de identificar tendências que podem ajudar no resultado financeiro de uma empresa.

### **3. O que se entende por Big Data?**

O termo “big data” segundo [6], refere-se a dados que apresentam uma grande dimensão, velocidade e/ou complexidade, que são difíceis ou impossíveis de processar usando métodos tradicionais. Embora o ato de aceder e armazenar grandes quantidades de informações para análise, exista há muito tempo, o conceito de big data obteve uma maior utilização no início dos anos 2000, quando o analista do setor Doug Laney articulou a definição agora dominante de big data como os três V's:

- **Volume:** as organizações adquirem dados de uma variedade de fontes, incluindo transações comerciais, dispositivos inteligentes (IoT), equipamentos industriais, vídeos, mídia social e muito mais. No passado, armazená-lo seria um problema - mas o armazenamento mais barato em plataformas como data lakes e Hadoop aliviou o fardo.
- **Velocidade:** Com o crescimento da Internet das Coisas, os dados chegam às empresas a uma velocidade sem precedentes e devem ser tratados em tempo hábil. Etiquetas RFID, sensores e medidores inteligentes levam à necessidade de lidar com essas correntes de dados quase em tempo real.
- **Variedade:** os dados vêm em todos os tipos de formatos - de dados numéricos estruturados em bancos de dados tradicionais a documentos de texto não estruturados, e-mails, vídeos, áudios, dados de cotações da bolsa e transações financeiras e cotações da bolsa e transações financeiras.

#### 4. O que é um modelo?

Um Modelo é uma ferramenta que permite demonstrar como serão construídas as estruturas de dados que darão suporte aos processos de negócio, como estes dados estarão organizados e quais os relacionamentos que pretendemos estabelecer entre eles.

A abordagem que se dispensa ao assunto normalmente atende a três perspectivas:

- Modelagem Conceitual: usada como representação de alto nível e considerada exclusivamente o ponto de vista do utilizador criador dos dados;
- Modelagem Lógica: agrega mais alguns detalhes de implementação.
- Modelagem Física: demonstra como os dados são fisicamente armazenados.

Quanto ao objetivo, podemos identificar as seguintes variações: Modelagem de dados entidade-relacionamento (leitura, construção e validação dos modelos); Modelagem de relacionamentos complexos (grupos de dados lógicos e ciclo de vida das entidades); Modelagem de dados corporativa; Modelagem de dados distribuídos (cliente/servidor); Modelagem e reengenharia de dados legados e modelagem de dados para Data Warehouse [3].

#### BIBLIOGRAFIA

- [1] Ciência de Dados: O que é, conceito e definição - <https://www.cetax.com.br/blog/data-science-ou-ciencia-de-dados/>
- [2] Ciência de dados - [https://pt.wikipedia.org/wiki/Ci%C3%Aancia\\_de\\_dados](https://pt.wikipedia.org/wiki/Ci%C3%Aancia_de_dados)
- [3] Modelagem de dados - [https://pt.wikipedia.org/wiki/Modelagem\\_de\\_dados](https://pt.wikipedia.org/wiki/Modelagem_de_dados)
- [4] Data Science: O que é, conceito e definição - <https://www.cetax.com.br/blog/data-science/>
- [5] Cientistas de Dados - [https://www.sas.com/pt\\_br/insights/analytics/cientistas-de-dados.html](https://www.sas.com/pt_br/insights/analytics/cientistas-de-dados.html)
- [6] Big Data: What it is and why it matters - [https://www.sas.com/pt\\_br/insights/big-data/what-is-big-data.html](https://www.sas.com/pt_br/insights/big-data/what-is-big-data.html)