Introdução a Ciência de Dados 3.0 com a Data Science Academy

Dúvidas: <https://www.datascienceacademy.com.br/faq>

Suporte: suporte@datascienceacademy.com.br

**O que é a Ciência de Dados?**

É o termo usado para definir a extração de insihts de dados que são coletados de várias fontes. Utilizando de várias técnicas, incluindo modelagem preditiva (Machine Learning), a Ciência de Dados ajuda a analisar e interpretar grandes quantidades de dados.

Ciência de Dados é o proceso para extrair informações valiosas a partir de dados. Como estamos vivendo na era do Big Data, a Ciência de Dados está se tornando um campo cada vez mais importante para explorar e processar grandes volumes de dados, com diferentes formatos, gerados a partir de várias fontes e em diferentes velocidades.

Ciência de Dados é uma grande disciplina em si e consiste em conjuntos e habilidades especializadas, ais como: estatística, matemática, programação, computação e conhecimentos de negócios, além de técnicas e teorias, como a análise preditiva, modelagem, mineração de dados e visualização de dados.

O principal objetivo da Ciência da Dados é extrair e interpretar os dados de forma eficaz e apresentá-los em uma linguagem simples para os usuários finais e tomadores de decisão.

**Fatos sobre a Ciência de Dados:**

1. Usando a analize de Big Data a Netflix economiza U$1 bilhão por ano em retenção de clientes;
2. O mercado de análises de Big Data em saúde pode valor U$67,82 bilhoes em 2025;
3. Em 2020, o mercado de análise de Big Data cresceu 14%;
4. O número de dispositivos IoT pode aumentar para 41,6 bilhões em 2025;
5. As vagas de empregos para funções em Ciência de Dados alcançaram cerca de 2,7 milhões em 2020 em todo mundo.
6. As vagas em Ciência de Dados aumentarão em cerca de 28% até 2026;

Esses fatos demonstram que precisaremos mais do que nunca de técnicas, ferramentas, procedimentos e processos para analisar os dados, extraindo valor e ajudando as empresas a aumentar lucro, reduzir custo e sobreviver em um mercado cada vez mais competitivo e orientada a dados (data-driven).

**Áreas de Conhecimento:**

A ciência de dados é uma área multidisciplinar, sendo necessário abstração de várias matérias. Pode ser dividida em 3 grandes pilares, Matemática e Estatística, Ciência da Computação e Conhecimento das áreas de negócio.

Matemática e Estatística: Fornecem técnicas e proceidmentos de cálculos, análise, correlação, transformação, limpeza e interpretação dos dados.

Ciência da Computação: Fornece ferramentas usadas no processo de análise, tal como programação de computadores, armazenamento e processamento paralelo e otimização do tempo de execução dos processos de análise.

Conhecimento das áreas de negócio: O objetivo da Ciência de Dados é resolver problemas de negócio e o conhecimento de domínio (área de negócio) é fundamental para saber interpretar os dados e os resultados do processo de análise.

**Em essência, a Ciência de Dados envolve o uso de métodos automatizados (Ciência da Computação) para analisar (Matemática e Estatística) enormes quantidades de dados a fim de extrair conhecimento (Área de Negócios) a partir dos dados.**

O dado bruto não tem valor para tomada de decisão, o grande lance está na ciência que aplica técnicas e conhecimento de diversas áreas para tratar, manipular, extrair, enfim os dados.

Ciência de Dados é a exploração e análise de todos os dados disponíveis, sejam eles estruturados ou não, com o objetivo de desenvolver compreensão, extrair conhecimento e formular ações que gerem resultados.

Diagrama

Descrição gerada automaticamente

**Dados e Tomada de Decisão:**

Todo projeto de ciência de dados começa com a definição do problema de negócio.

Exemplo:

Problema identificado: Clientes cancelam seus planos de telefonia com frequência de 2 a 3 meses após assinatura do plano.

1. Dados e Análise: Dados coletados e analisados, levaram à conclusão que existe uma tendência entre os clientes de operadoras de telefonia em trocar de operadora após o terceiro contato com atendimento ao cliente.
2. Decisão: Os executivos da empresa, baseados em dados, decidiram reve todos os processos de atendimento ao cliente.
3. Ação: A empresa contratou uma consultoria para fornecer treinamento especializado de atendimento ao cliente, passou a monitorar as ligações, criou programas de incentivo aos melhores operadores, etc...

O principal objetivo é extrair informações dos dados e transformá-las em conhecimento que possam ser usadas para a tomada de decisões.

**O processo Data-Driven:**

Processo orientado a dados –

Dados >> Ação

O que aconteceu? (Identificar o problema de negócio)

Por que aconteceu? (Através de ferramentas e técnicas, descobrimos a origem ou causas)

Acontecerá novamente? (Machine Learning)

O que deve ser feito? (É decisão dos executivos superiores, porém podemos fazer sugestões enquanto analistas envolvidos nas descobertas...)

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

**O que são dados?**

São fatos individuais, estatísticas ou medidas, que são coletados por meio de observação e/ou medição. Em um sentido mais técnico, os dados são um conjunto de valores de variáveis qualitativas ou quantitativvas sobre eventos, pessoas ou objetos. Embora o termos “dados” e “informações” sejam frequentemente usados de forma intercambiável, esses termos têm significados distintos.

**De onde vem os dados?**

Eles estão presentes em toda forma de representação quantitativa ou qualitativa que se julgue relevante para determinado fim. Uma anotação em uma folha de papel, uma planilha de excel, uma imagem, vídeo, audio...

A partir do que se busca, realizamos a coleta direcionada de acordo com a definição do problema de negócio, porque tudo gera custo associado, no caso a coleta e armazenamento também geram custos.

**O que os dados representam?**

Dados que representam Entidades;

Dados que representam Características;

Dados que representam Ambientes;

Dados que representam Eventos;

Dados que representam Comportamentos;

Dados que representam Resultados;

Dados que representam Observação;

Dados que representam uma Medição;

Dados que representam uma Suposição;

**Dados e Informação**

Os dados pdem ser qualquer caracter, texto, palavras, números, imagens, som ou vídeo e, se não colocados em um contexto, significa pouco ou nada para um ser humano. No entanto, o dado aplicado a um contexto, gera a informação que é útil quando formatada de uma maneira geral, permitindo que seja entendida por um ser humano. Então para os gerentes e tomadores de decisões, os dados em si não são úteis, eles precisam que os analistas e cientistas de dados transformem esses dados em informação útil para tomada de decisão.

**O que são datasets**

Conjuntos de dados;

Coleções de observações;

Cada observação é tipicamente chamada de registro;

Cada registro tem um conjunto de atributos que apontam características, comportamentos ou resultados;

Datasets podem ter 1 registo e 1 atributo e não há limite máximo.

**O Big Data**

São dados como frequência de compra, consmo médio, número de curtidas em posts, pesquisas, etc... Multiplique isso bilhões de vezes e você começa a perceber o que é o Biga Data. É uma coleção de conjuntos de dados, grandes e complexos, que não podem ser processados por bancoss de dados ou aplicações de processamentos tradicionais.

Definido por 4 V’s:

Volume – quantidade de dados;

Variedade – dados estruturados, não estruturados, semi-estruturados;

Velocidade – os dados são gerados cada vez a mais velocidade;

Veracidade – os dados que compoem o big data são dados reais.

Pode ser composto por dados de diferentes fontes e diferentes formatos:

Uma imagem contendo Tabela

Descrição gerada automaticamente

**Relação Ciência de Dados e Big Data**

Big Data é o combustível para a ciência de dados. A grande oportunidade do Big Data:

Interface gráfica do usuário, Site

Descrição gerada automaticamente

“Nós tendemos a superestimar o efeito de uma nova tecnologia no curto prazo e subestimas o efeito no longo prazo.” – Roy Amara

**O que é e como aprender Big Data Analytics?**

Big Data Analytics é a aplicação da Ciência de Dados para a análise de Big Data. Com Big Data Analytics o objetivo é resolver problemas de negócio, através de análise de dados com alto volume, variedade e gerados em alta velocidade.