Módulo 1 – Aula 01 – INTRODUÇÃO AO ECOSSISTEMA

BIGA DATA (precisa ter alto volume (qtd física, 40 zetabytes até o final do ano), velocidade e variedade, veracidade e valor)

O que é?

Principais V’s

Aplicação prática

Quantidade de dados

MACHINE LEARNING

Conceitos, tipos de aprendizagem, deep learning...

Benefícios: tomada de decisão (orientada por dados), regressão (prever resultado numérico usando uma ou mais variáveis de entrada), clustering, detecção de anomalias...

Aprendizagem (supervisionada)

Conjunto prévio entrada/saída de dados; utilizado em regressão e classificação.

Overfitting – modelo se ajusta demais aos dados (decora ao invés de aprender)

Underfitting – modelo é muito simples em relação aos dados

DATA MINING

KDD (descobrir conhecimento em conjunto de dados), é um processo de várias etapas...

Data Mining

Qualidade dos dados

Desafios e limitações

DATA ANALYTICS E DATA SCIENCE

Só faz análise de dados se quiser resolver um problema

BUSINESS INTELLIGENCE

Aplicações, infra, ferramentas, e boas práticas que possibilita a otimização de decisões e performances.

K-MEANS

Baseado em aprendizado não supervisionado

ETL é um processo, e não uma ferramenta