



WORKSHOP
Digital Data Analytics
Áreas de estudo

Jorge da Costa Ferreira



WORKSHOP

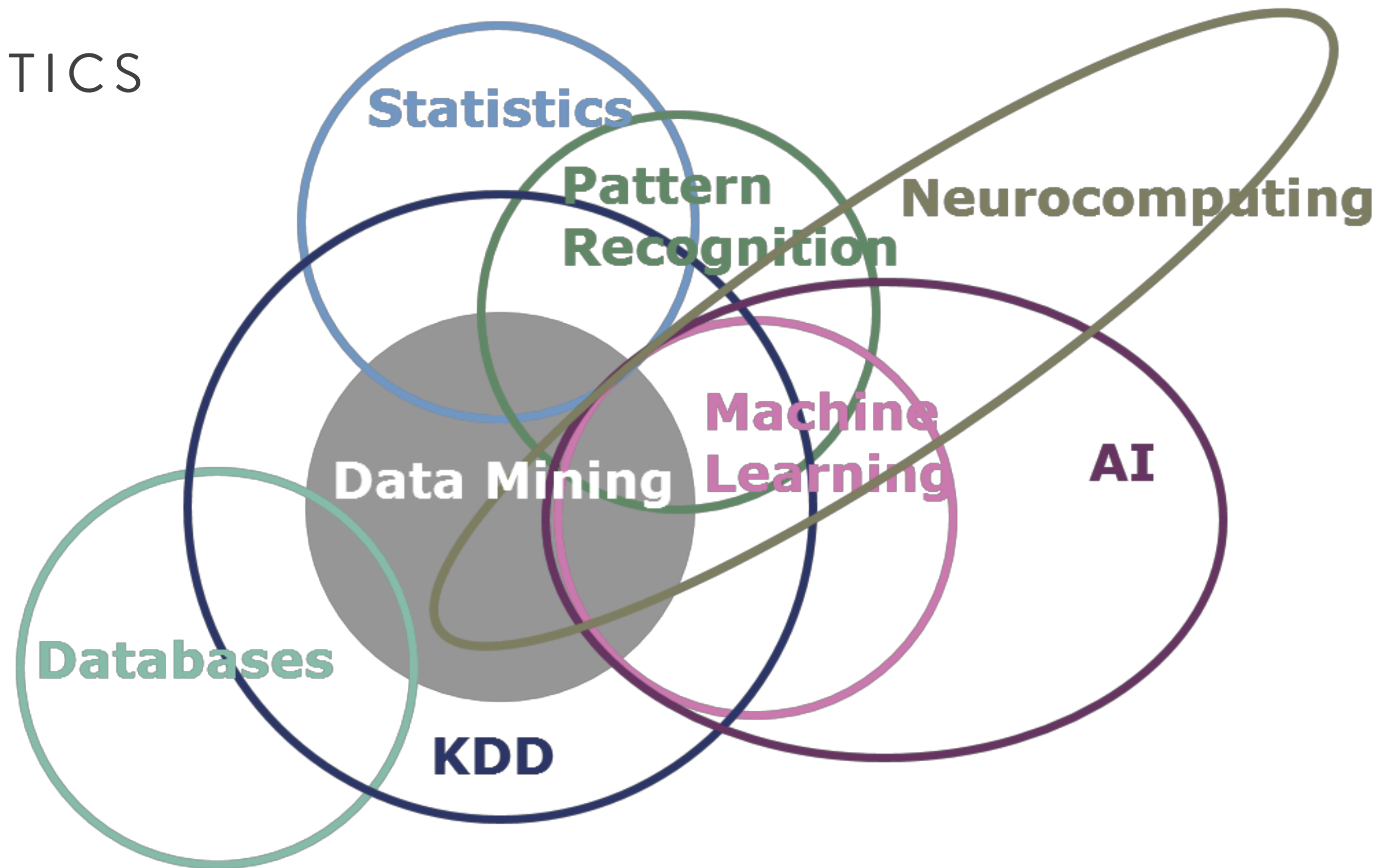
Digital Data Analytics

MAS QUAIS SÃO AS FERRAMENTAS UTILIZADAS PARA REPRODUZIR ESTAS ANÁLISES?

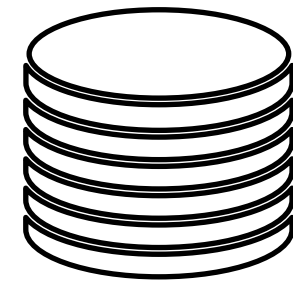
FERRAMENTAS DO DATA ANALYTICS

O data analytics é composto por diversas áreas de aplicação que vão desde a **estatística**, as **bases de dados**, o **data mining**, o **machine learning**, a **inteligência artificial**, etc.

Muitas destas áreas se sobrepõem e têm propósitos semelhantes e/ou complementares.



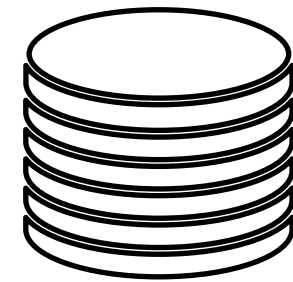
BASES DE DADOS



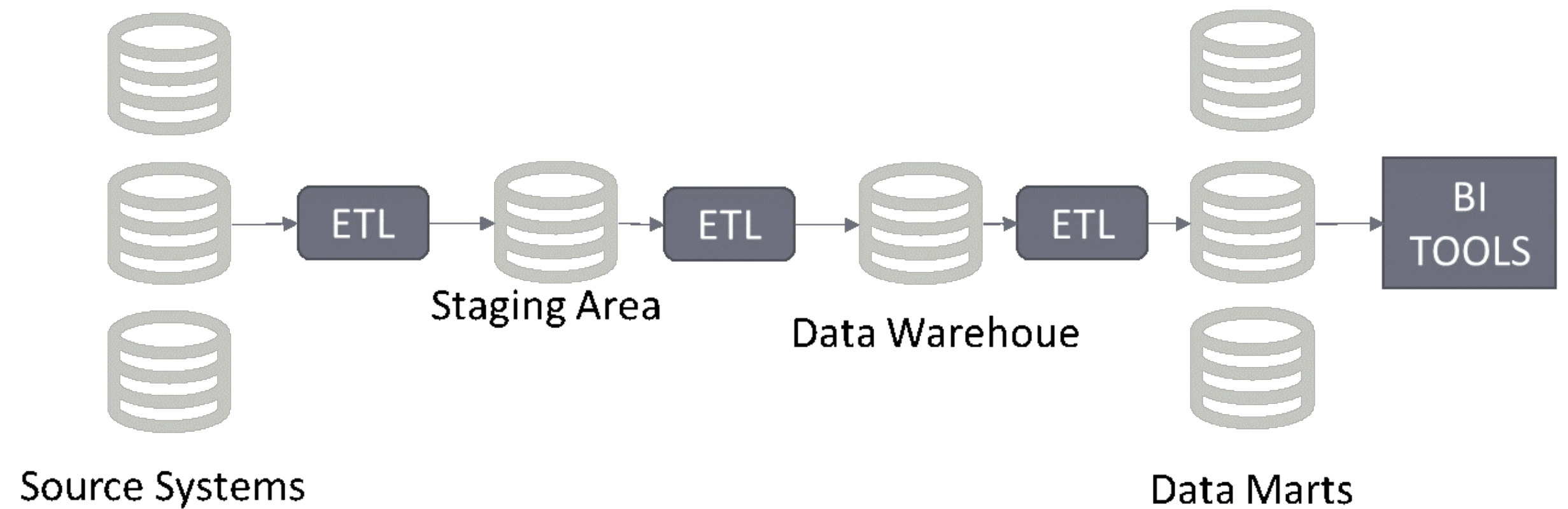
- Coleção de informação, organizada de forma a ser facilmente acessível, gerida e atualizada
- Os dados podem ser inseridos, consultados, atualizados e apagados
- Tipicamente (SQL) os dados são organizados em linhas, colunas e tabelas, de forma indexada para que seja fácil a consulta:
- Existem no entanto novas vertentes de base de dados não relacionais NoSQL
- Existem bases de dados orientadas a transações e orientadas a análise

Estima-se que a cada 20 meses a quantidade de dados guardada em todas as bases de dados do mundo duplique (Witten et al., 2011)

BASES DE DADOS

**Exemplo:**

O Data Mart de Marketing relaciona o número de vendas com o preço dos itens, enquanto o Data Mart de Logística relaciona o número de vendas com o número de transportes



No entanto, para se construir processos de data analytics os dados podem estar alojados em qualquer formato e em diversos volumes. Desde um data lake até uma folha de cálculo excel.

DATA MINING



Ciência multi-disciplinar que recorre a diferentes técnicas para descobrir propriedades nos dados, desde **machine learning**, **estatística** ou, simplesmente, olhar para os dados.

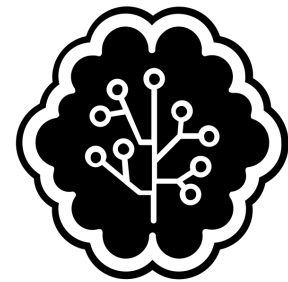


Muito orientada à construção de **modelos probabilísticos**.

Serve propósitos idênticos ao machine learning com abordagens mais tradicionais.



MACHINE LEARNING

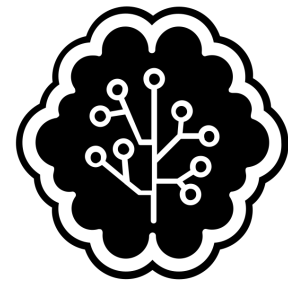


Campo da área da inteligência artificial, que visa dotar as máquinas a **capacidade de tomar decisões**, tornando-as agentes com “**inteligência**”.

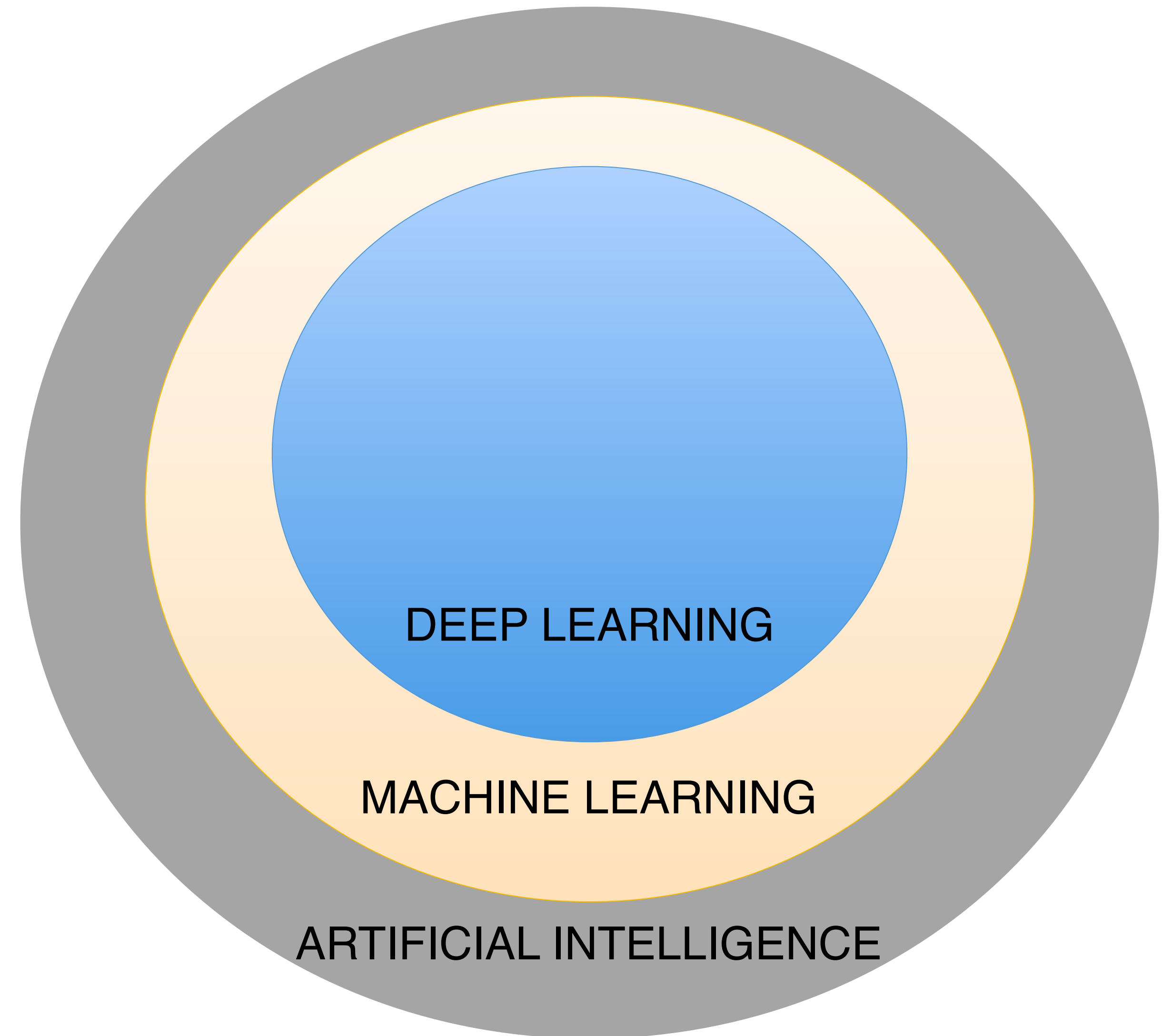
Baseado na modelação de algoritmos que resolvem classes de problemas através da extração de conhecimento dos dados.



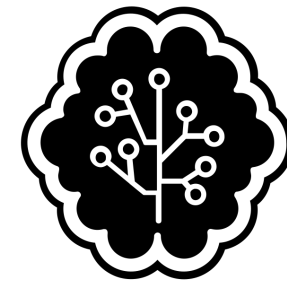
DEEP LEARNING



- Um subset do Machine Learning
- Utiliza Redes Neurais
- As variáveis relevantes são extraídas através dos dados
- Requer um enorme volume de dados
- Necessita de um grande poder computacional



DEEP LEARNING

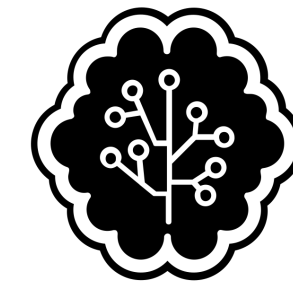


- Um subset do Machine Learning
- Utiliza Redes Neurais
- As variáveis relevantes são extraídas através dos dados
- Requer um enorme volume de dados
- Necessita de um grande poder computacional

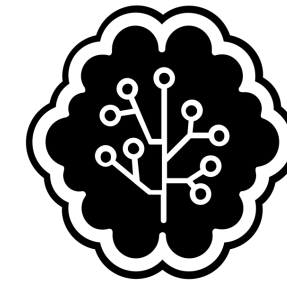




IS DEEP LEARNING OVERHYPED?



IS DEEP LEARNING OVERHYPED?



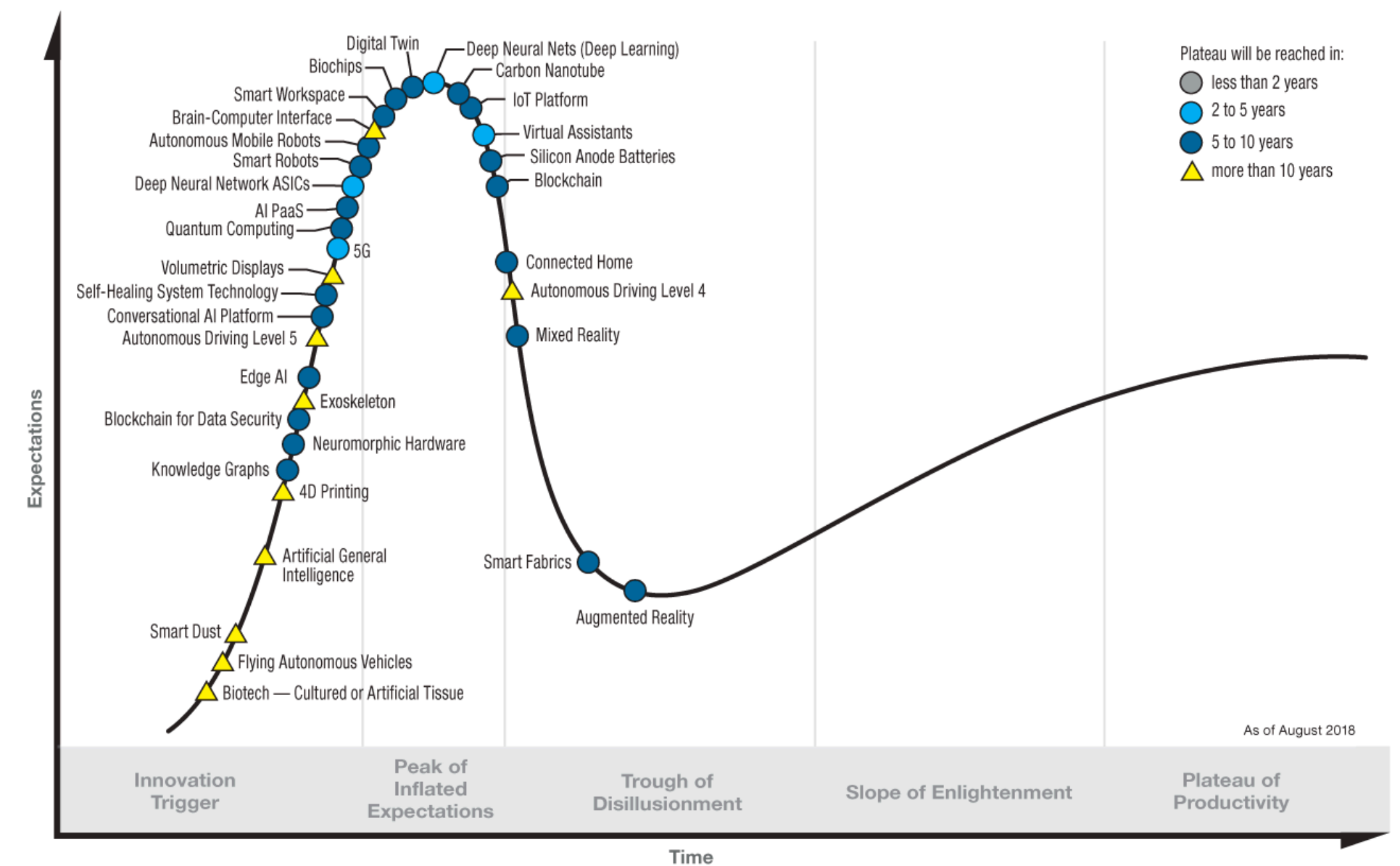
2017



Source: Gartner (July 2016)

2018

Hype Cycle for Emerging Technologies, 2018



gartner.com/SmarterWithGartner

Source: Gartner (August 2018)
© 2018 Gartner, Inc. and/or its affiliates. All rights reserved.

Gartner

WORKSHOP
Digital Data Analytics
Áreas de estudo