# MINERAÇÃO DE DADOS

Prof. Me. Napoleão Póvoa Ribeiro Filho





### **DIFERENTES NOMENCLATURAS**

- Inteligência artificial
- Mineração de dados
- Aprendizado de máquina



# INTELIGÊNCIA ARTIFICIAL (IA)

A lA envolve a criação de sistemas computacionais capazes de realizar tarefas que normalmente exigiriam inteligência humana, como reconhecimento de padrões, raciocínio lógico e tomada de decisão.

Russell, S., & Norvig, P. (2021). Artificial Intelligence: A Modern Approach. Pearson.



# MINERAÇÃO DE DADOS

- É o processo de descoberta de padrões úteis, informações relevantes e relações ocultas em grandes volumes de dados.
- Essa área combina estatística, aprendizado de máquina e bancos de dados para transformar dados brutos em conhecimento útil.

Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques. Elsevier.



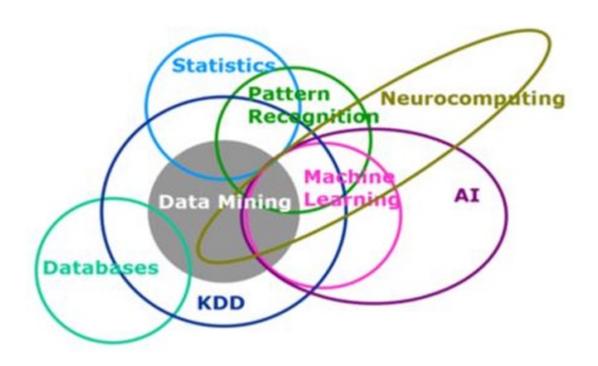
# APRENDIZADO DE MÁQUINA (ML)

- É um subcampo da Inteligência Artificial que estuda algoritmos que permitem aos computadores aprender a partir de dados, sem serem explicitamente programados.
- Um sistema aprende se seu desempenho em uma tarefa melhora com a experiência adquirida.

Mitchell, T. M. (1997). Machine Learning. McGraw-Hill

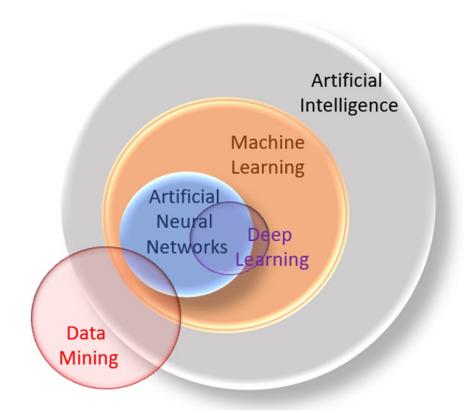


# RELAÇÃO ENTRE IA x MD x ML





# RELAÇÃO ENTRE IA x MD x ML





#### IA x ML x MD

- IA é o conceito mais amplo, englobando diversas técnicas para simular inteligência humana.
- Aprendizado de Máquina é um subcampo da IA que permite que máquinas aprendam padrões a partir de dados.
- Mineração de Dados utiliza técnicas de Aprendizado de Máquina e estatística para descobrir padrões úteis em grandes bases de dados.



#### PARADIGMAS DE APARENDIZAGEM

- Classificam como os algoritmos aprendem a partir dos dados.
- Existem três paradigmas principais: aprendizado supervisionado, aprendizado não supervisionado e aprendizado por reforço.
- Contudo, existem abordagens híbridas: aprendizado semi-supervisionado e aprendizado por autoaprendizagem



#### APRENDIZADO SUPERVISIONADO

- O modelo é treinado com um conjunto de dados rotulados, onde cada entrada tem uma saída esperada.
- O objetivo é aprender um mapeamento entre entradas e saídas para generalizar novos exemplos.



#### APRENDIZADO SUPERVISIONADO

- Classificação: Um modelo treinado para classificar e-mails como spam ou não spam.
- Regressão: Predição do preço de imóveis com base em variáveis como tamanho, localização e número de quartos.



## APRENDIZADO NÃO-SUPERVISIONADO

 No aprendizado n\u00e3o supervisionado, o modelo recebe apenas os dados de entrada, sem r\u00f3tulos, e deve identificar padr\u00f3es, agrupamentos ou estruturas ocultas.



## APRENDIZADO NÃO-SUPERVISIONADO

- Agrupamento (Clustering): Um algoritmo como k-means pode segmentar clientes de um e-commerce com base em comportamento de compra, sem informações prévias sobre grupos.
- Redução de Dimensionalidade: Técnicas como PCA (Principal Component Analysis) são usadas para reduzir a complexidade dos dados.



## APRENDIZADO POR REFORÇO

- No aprendizado por reforço, um agente interage com um ambiente, toma ações e recebe recompensas ou penalidades.
- O objetivo é aprender uma política ótima para maximizar as recompensas ao longo do tempo.

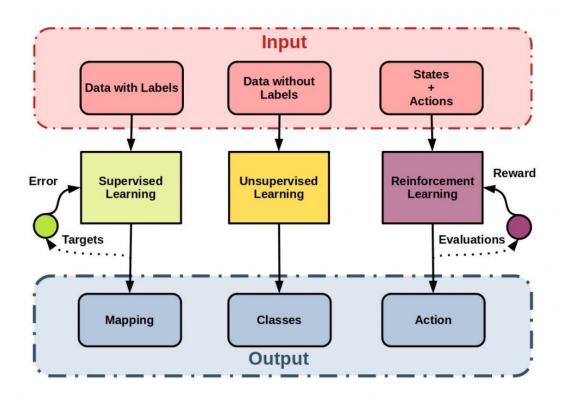


## APRENDIZADO POR REFORÇO

 Jogos e Robótica: Algoritmos de Deep Q-Learning treinam agentes para jogar xadrez ou dirigir carros autônomos aprendendo por tentativa e erro.



### PRINCIPAIS PARADIGMAS





### APRENDIZADO SEMI-SUPERVISIONADO

 O aprendizado semi-supervisionado combina aprendizado supervisionado e não supervisionado, utilizando um pequeno conjunto de dados rotulados e um grande conjunto de dados não rotulados para treinar o modelo.



#### APRENDIZADO SEMI-SUPERVISIONADO

 Reconhecimento de imagens: Em um sistema de reconhecimento facial, apenas uma parte das imagens tem rótulos de identidade, e o modelo aprende a generalizar com as demais.



#### APRENDIZADO POR AUTOAPRENDIZAGEM

- Esse paradigma é uma variação do aprendizado supervisionado, onde o próprio modelo gera rótulos a partir de padrões nos dados.
- Modelos de linguagem (LLMs): O GPT e o BERT são treinados prevendo palavras ocultas em frases sem precisar de rótulos humanos explícitos.



