

# INTRODUÇÃO À CIENCIA DE DADOS

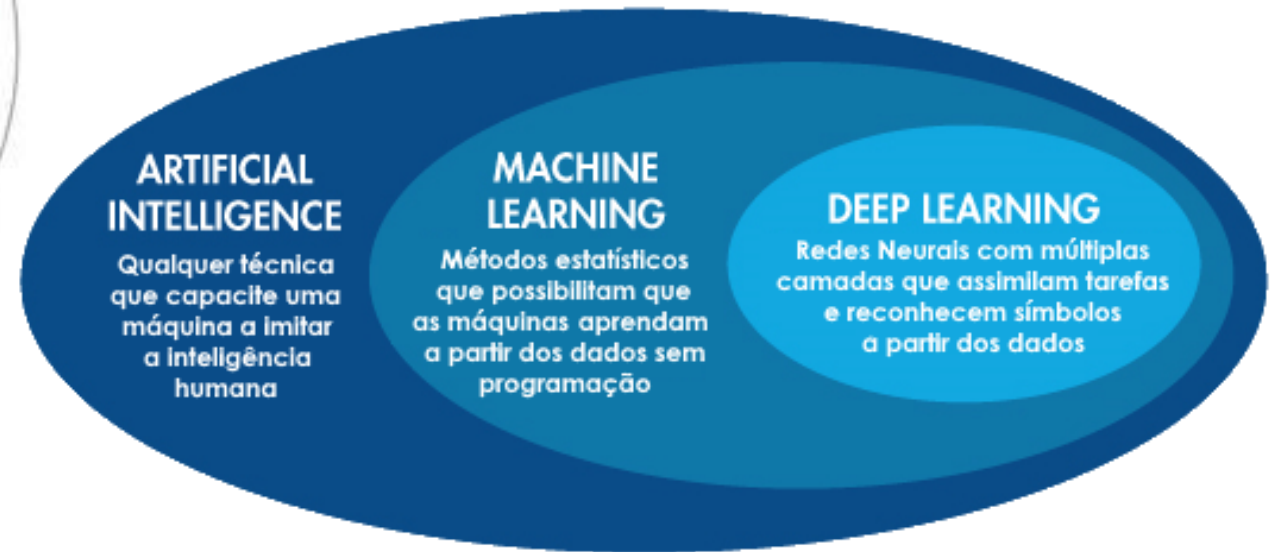


# Ciência de Dados e Machine Learning

**Todo projeto de Machine Learning é um projeto de Ciência de Dados.**

**Mas nem todo projeto de Ciência de Dados precisa envolver Machine Learning.**

# Ciência de Dados e Machine Learning



# Ciência de Dados e Machine Learning

## Inteligência + Artificial

- Vocês já refletiram sobre a definição destas duas palavras?
- **Inteligência:** habilidades mentais de conhecer, compreender, aprender, resolver problemas e adaptar-se.
  - *inter*: “entre” + *legere*: “escolha” => entre escolhas.
- **Artificial:** produzido pela mão do homem, não pela natureza, geralmente como uma cópia de algo natural.

***Habilidades mentais desenvolvidas de forma não natural pela mão do homem.***

***Cópias de habilidades mentais humanas criadas de forma não natural pelo homem.***





# Ciência de Dados e Machine Learning

## Definições e Objetivo da IA

- **Definição:** “Capacidade de um sistema de interpretar corretamente dados externos (vindos do ambiente), aprender com esses dados e usá-los para atingir objetivos específicos por meio de adaptação flexível” (Andreas Kaplan).
- **Definição:** “Ciência e engenharia de produzir máquinas inteligentes” (John McCarthy).
- **Objetivo:** Criar máquinas que *imitem* nossa capacidade mental para uma determinada tarefa.
- Essa *imitação* é apenas uma aproximação pois ainda não conseguimos criar matéria viva.
- É por isso que em IA fala-se da criação de máquinas que são **modelos aproximados** de nossa capacidade de aprender, raciocinar, enxergar, falar, ouvir, etc.

# Ciência de Dados e Machine Learning

Inteligência Artificial (IA) é uma área ampla que engloba várias aplicações (ou subáreas ou ainda objetivos) tais como:

- (i) processamento de linguagem natural
- (ii) representação do conhecimento
- (iii) raciocínio automatizado
- (iv) visão computacional
- (v) Robótica
- (vi) **Aprendizado de Máquina** que por sua vez engloba redes neurais artificiais, etc

# Ciência de Dados e Machine Learning

**Aprendizado** é uma das propriedades mais importantes da inteligência humana. É a capacidade de se adaptar, modificar e melhorar seu comportamento e suas respostas.

Exemplo: Um problema matemático pode ser solucionado (i) caso tenha sido resolvido antes e sua solução aprendida ou (ii) caso ele nunca tenha sido visto antes mas um indivíduo com inteligência imagina possíveis maneiras de como resolve-lo baseado no que ele sabe (conhecimento).

Objetivo do **Machine Learning**: Tentativa de reproduzir o processo de aprendizado de seres humanos em máquinas.

# Ciência de Dados e Machine Learning

## O que é Machine Learning?

- Machine Learning ou Aprendizado de Máquina é um subcampo da Inteligência Artificial que permite dar aos computadores a habilidade de aprender sem que sejam explicitamente programados para isto.
- Usando algoritmos que aprendem a partir de dados, o aprendizado de máquina permite que os computadores encontrem padrões ocultos sem que sejam explicitamente programados para procurar algo específico.



# Ciência de Dados e Machine Learning

## O que é Machine Learning?

- Algoritmos de ML são orientados a dados, ou seja, eles aprendem automaticamente um modelo geral (i.e., generalizar) a partir de grandes volumes de dados.
- Exemplo: filtro de spam do gmail.

# Ciência de Dados e Machine Learning

## **Por que Machine Learning se tornou tão importante?**

- Vivemos na era da informação. Grandes volumes de dados disponíveis, impossibilitando sua análise por nós seres humanos.
- Porém, para modelos de ML isso não é um problema, quanto mais dados melhor será o aprendizado.
- Hoje em dia, dados são preciosíssimos e a extração de novas informações (úteis) vale ouro.
- Surgimento de recursos computacionais poderosos tais como GPUs e CPUs com múltiplos cores.
- Existência de frameworks e bibliotecas que facilitam o desenvolvimento de soluções com ML. Ex: Pandas, TensorFlow, Keras,...

# Ciência de Dados e Machine Learning

## Tipos de Aprendizagem

- Os seres humanos tem diferentes formas de aprendizado. Como você aprende? Qual a sua estratégia?
- Com as máquinas é a mesma coisa. O processo de aprendizado humano foi reproduzido nas máquinas.

# Ciência de Dados e Machine Learning

## Tipos de Aprendizagem



# Ciência de Dados e Machine Learning

## Aprendizagem Supervisionada



Qual dessas figuras representa um aprendizado supervisionado?



# Ciência de Dados e Machine Learning

## Aprendizagem Supervisionada

Os Algoritmos de aprendizado supervisionado fazem previsões com base em conjunto de exemplos (dados históricos).



# Ciência de Dados e Machine Learning

## Aprendizagem Supervisionada



# Ciência de Dados e Machine Learning

## Aprendizagem Supervisionada por Classificação

Os dados de treinamento, chamados de características (features) ou exemplos que alimentam o algoritmo de ML incluem as soluções desejadas, chamadas de rótulos (labels). Ex:

Dados históricos

Variável 1 (Idade)	Variável 2 (IMC)	Variável 3 (Pressão Arterial)	Label (Diabetes)
23	34	123	Sim
48	25	129	Não
37	39	145	Sim
39	18	134	Não
29	27	131	Não

Algoritmos de ML



Modelo Preditivo



Entramos com novos dados referentes às variáveis 1,2,3 e o modelo será capaz de prever a ocorrência de diabetes.

# Ciência de Dados e Machine Learning

## Aprendizagem Supervisionada por Classificação

Os dados de treinamento, chamados de características (features) ou exemplos que alimentam o algoritmo de ML incluem as soluções desejadas, chamadas de rótulos (labels). Ex:

Dados históricos

Variável 1 (Idade)	Variável 2 (IMC)	Variável 3 (Pressão Arterial)	Label (Diabetes)
23	34	123	Sim
48	25	129	Não
37	39	145	Sim
39	18	134	Não
29	27	131	Não

Algoritmos de ML



Modelo Preditivo



Entramos com novos dados referentes às variáveis 1,2,3 e o modelo será capaz de prever a ocorrência de diabetes.

# Ciência de Dados e Machine Learning

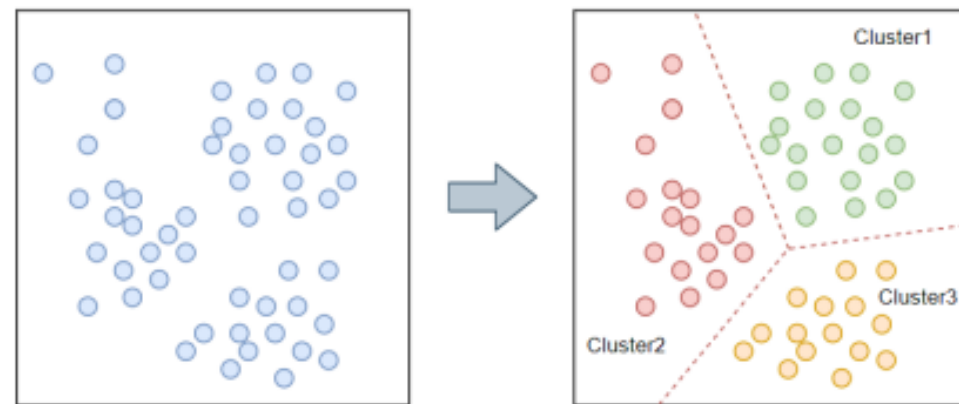
## Aprendizagem Não Supervisionada

As máquinas não são informadas sobre o que aprender. Elas só recebem os exemplos de treinamento.

Os algoritmos aprendem/descobrem padrões (ocultos) presentes nos dados de entrada sem a presença de rótulos.

Resumindo: Os modelos descobrem/aprendem padrões desconhecidos.

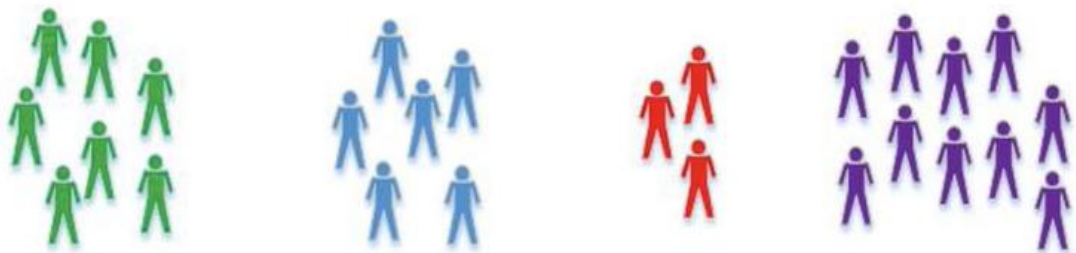
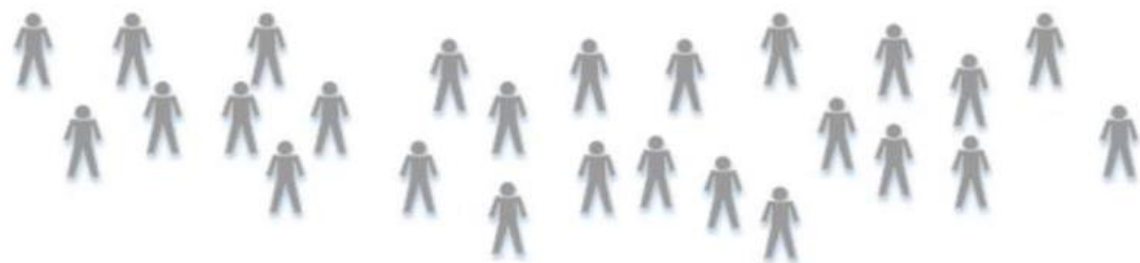
Utilizado em problemas de clusterização.





# Ciência de Dados e Machine Learning

## Aprendizagem Não Supervisionada



- É usado com dados que não possuem rótulos históricos, ou seja, não temos variáveis de saída para serem estimadas.
- O algoritmo deve descobrir a saída por si próprio. O objetivo é explorar os dados e encontrar alguma similaridade (padrão) neles.
- Estratégia muito usada em marketing para segmentação de clientes.

# Ciência de Dados e Machine Learning

## Aprendizagem Não Supervisionada

Dados Históricos

Variável 1 (Idade)	Variável 2 (Número de Compras)	Variável 3(média de compras)	
23	12	1200	A
48	3	800	C
37	4	750	B
39	18	400	A
29	9	680	A

Machine Learning



Modelo Preditivo



O modelo é capaz de agrupar os dados por similaridade encontrando grupos (segmentos) de clientes.

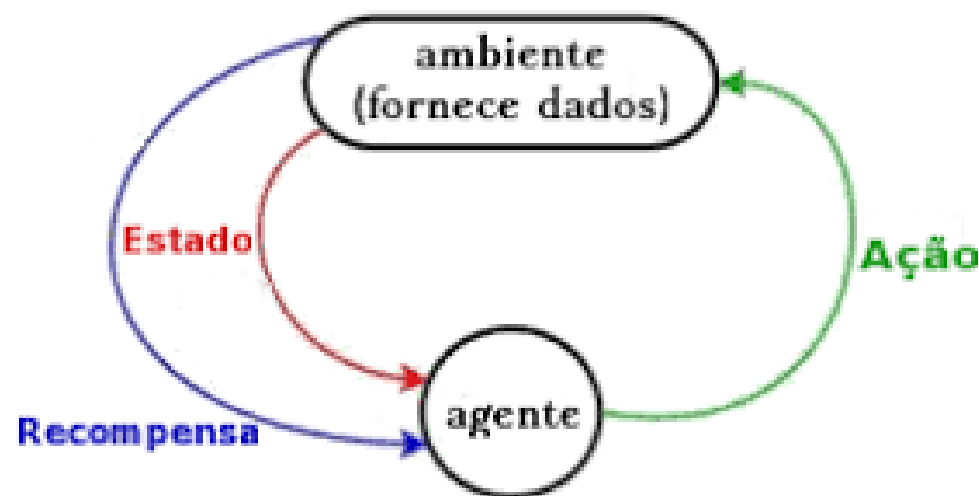
# Ciência de Dados e Machine Learning

## Aprendizagem por reforço

- O algoritmo de aprendizagem, chamado de agente nesse contexto, deve aprender como se comportar em um ambiente através de interações do tipo “tentativa e erro”.

O agente observa o estado do ambiente em que está inserido, seleciona e executa ações com base em políticas e recebe recompensas (ou reforço) em consequência das ações tomadas.

Estratégia muito usada em robótica e games e também usada por Robôs investidores.



# Ciência de Dados e Machine Learning

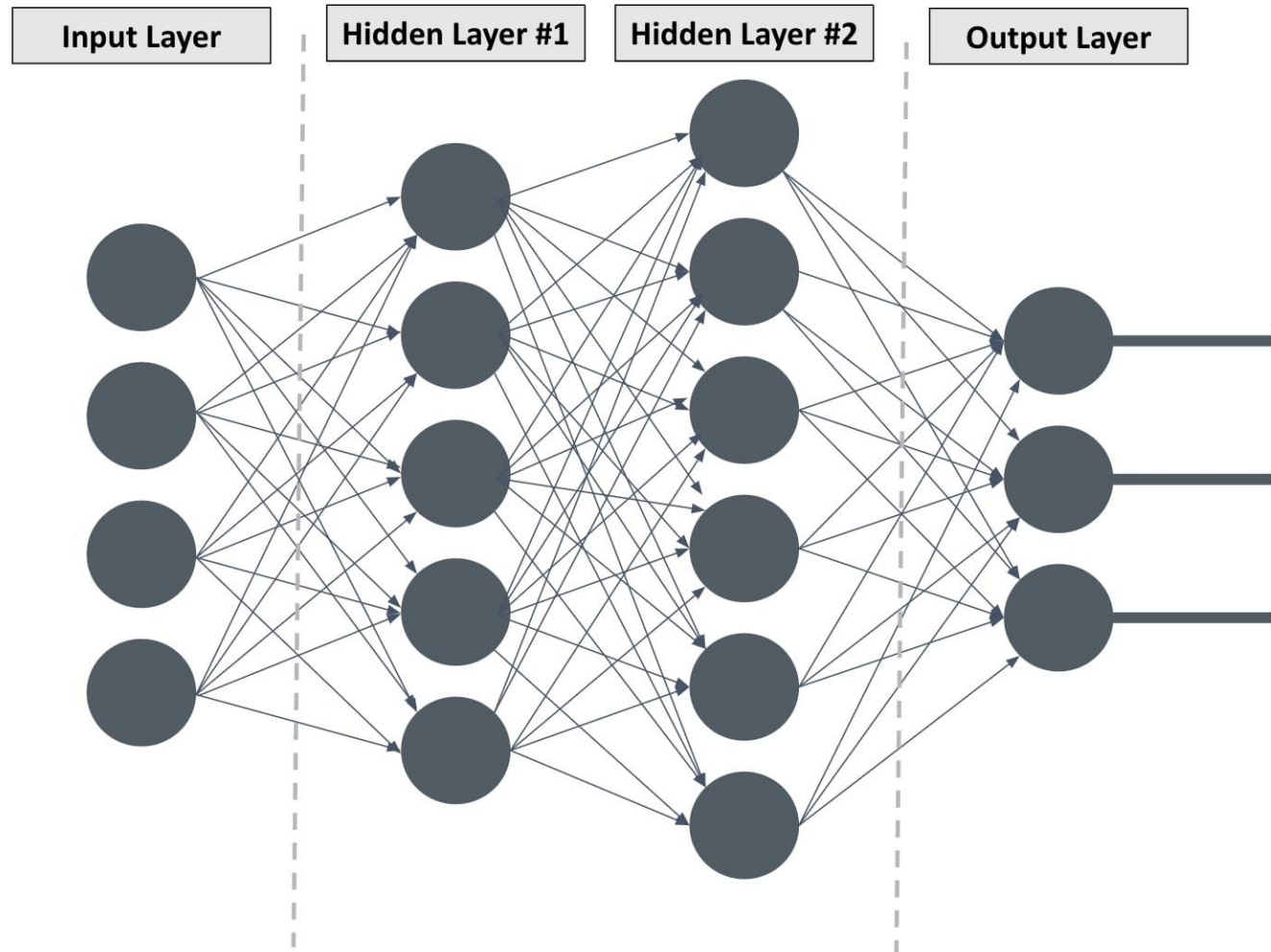
## Deep Learning

- **Deep learning** ou **Aprendizagem Profunda** é um tipo de **machine learning** que treina computadores para realizar tarefas como seres humanos, o que inclui reconhecimento de fala, identificação de imagem e previsões, sistemas de recomendação.

Em vez de organizar os dados para serem executados através de equações predefinidas, o deep learning configura parâmetros básicos sobre os dados e treina o computador para aprender sozinho através do reconhecimento padrões em várias camadas de processamento.

# Ciência de Dados e Machine Learning

## Deep Learning





# Ciência de Dados e Machine Learning

## Deep Learning

Nível de precisão altíssimo, requer muito poder computacional.  
Carros autônomos, drones autônomos (visão computacional)  
Assistentes virtuais (Siri, cortana) (processamento de linguagem natural)

A promessa do deep learning é concretizar sistemas preditivos que se difundem e se adaptam bem, melhoram continuamente à medida que novos dados são adicionados e são mais dinâmicos do que sistemas preditivos baseados em regras de negócios. Você não mais adapta um modelo, você o treina.

# Ciência de Dados e Machine Learning

**Obrigado!**