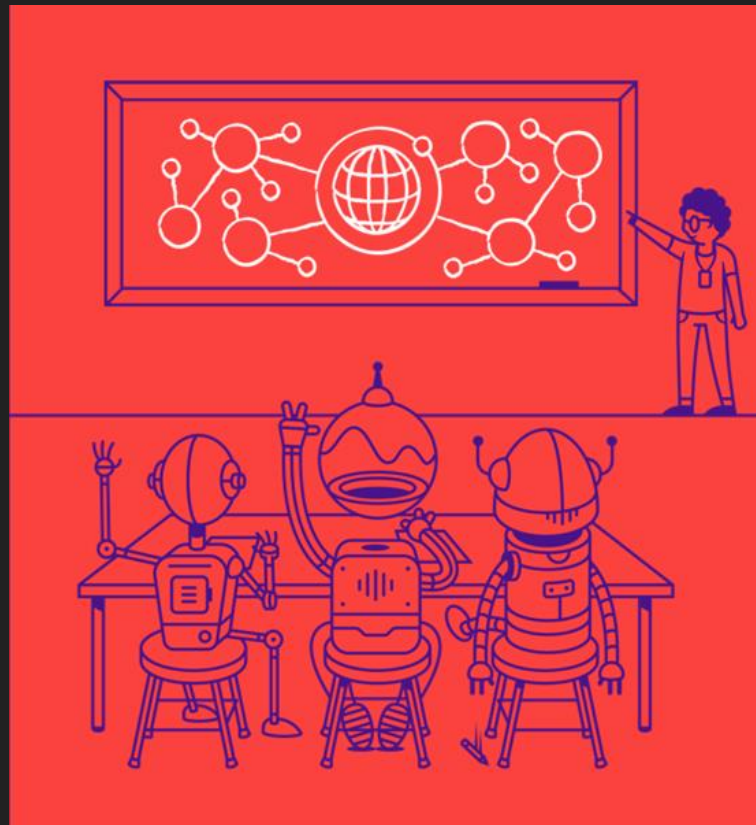


Data Science do ZERO

Capítulo 06 - Machine Learning
Conceitos.

O que é Machine Learning?

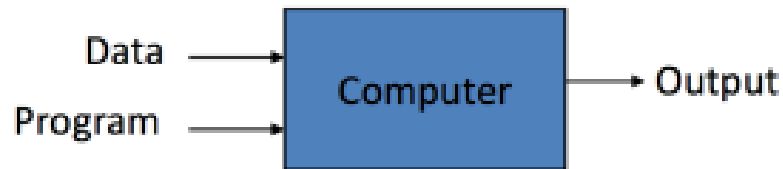
- Machine Learning significa **Aprendizado de Máquina**.
- Aprendizado de máquina é um subcampo da ciência da computação que evoluiu do estudo de reconhecimento de padrões e da teoria do aprendizado computacional em inteligência artificial (Wikipedia).
- Em 1959, Arthur Samuel definiu aprendizado de máquina como o "campo de estudo que dá aos computadores a habilidade de aprender sem serem explicitamente programados."



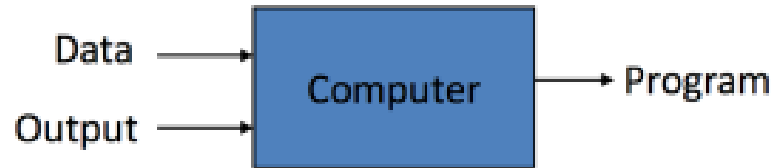
Porque Machine Learning?

- Por que usar Machine Learning?
 - Scalar sistemas atuais usando aprendizagem automática.
 - Desenvolvimento do sistemas é um gargalo, use Machine Learning para automatizar o desenvolvimento.
 - *A breakthrough in machine learning would be worth ten Microsofts.* (Bill Gates, Microsoft)

Traditional Programming

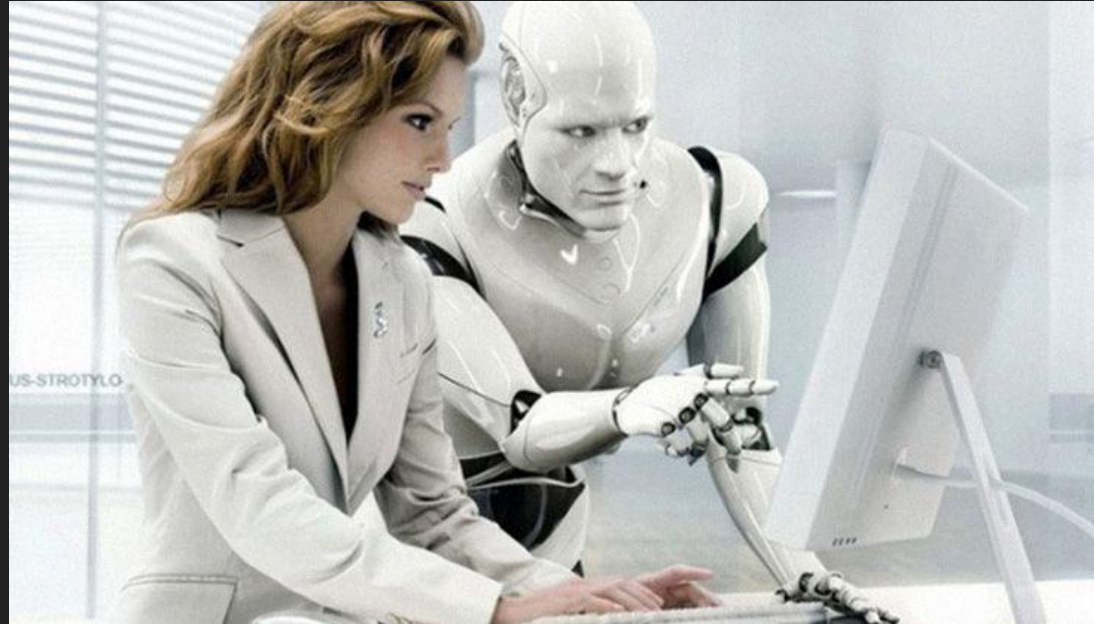


Machine Learning



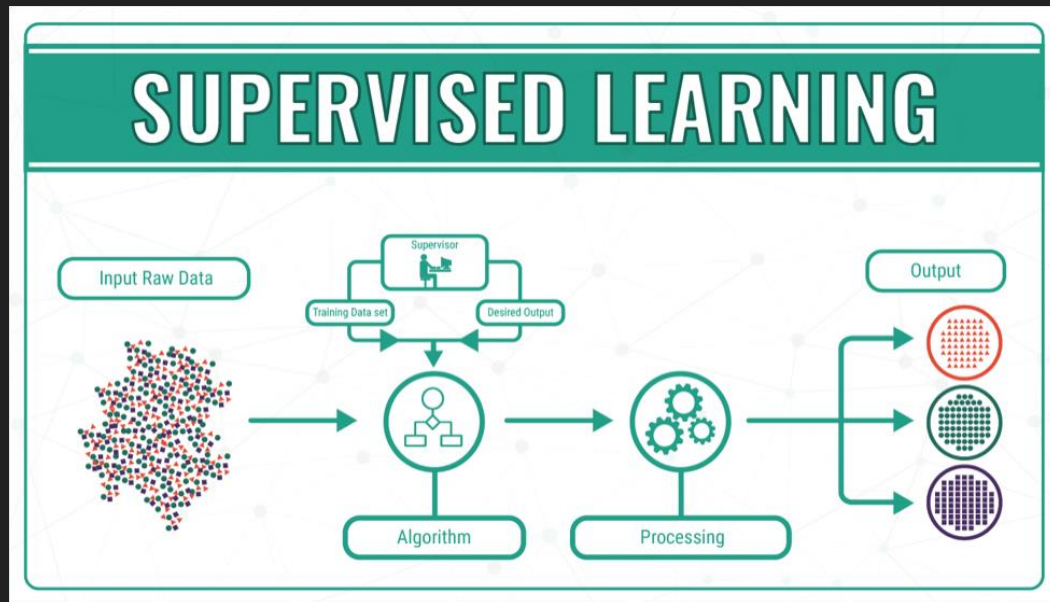
Tipos de Aprendizado

- Os tipos de aprendizado de máquina determinam o modo de funcionamento dos algoritmos.
- Existem alguns tipos de aprendizado dentre eles os mais comuns são:
 - Aprendizado Supervisionado
 - Aprendizado Não Supervisionado
 - Aprendizado Semi Supervisionado



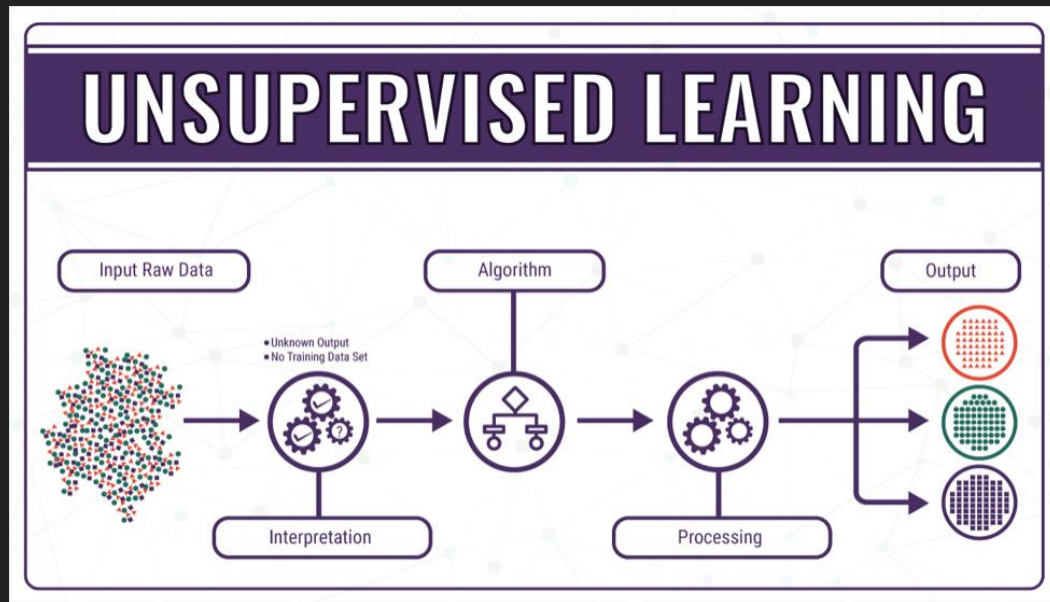
Tipos de Aprendizado

- O **Aprendizado Supervisionado** significa que há uma supervisão dos dados que serão usados pelo algoritmo para treino.
- A supervisão é dada pela rotulagem dos dados de exemplos.
- Normalmente essa supervisão é feita por humanos.
- Um exemplo seria onde temos exemplos de e-mails que são '*spam*' e '*nãospam*'. A rotulagem desses dados é dada por uma tarefa de supervisão.
- A principal desvantagem é o custo de rotulagem dos dados



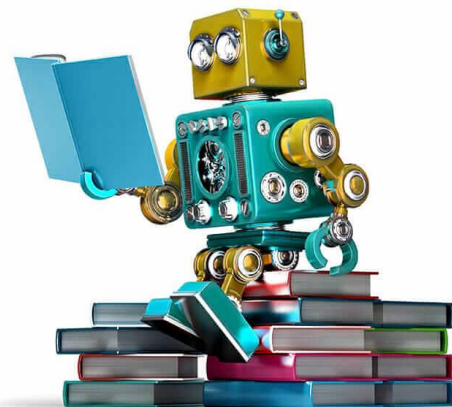
Tipos de Aprendizado

- O **Aprendizado Não Supervisionado** significa que não há uma supervisão dos dados que serão usados pelo algoritmo para treino.
- Com essa abordagem o algoritmo não recebe dados rotulados.
- Um exemplo seria um algoritmo que faz agrupamento de dados.
- A principal desvantagem é a forma de validar a performance do algoritmo.

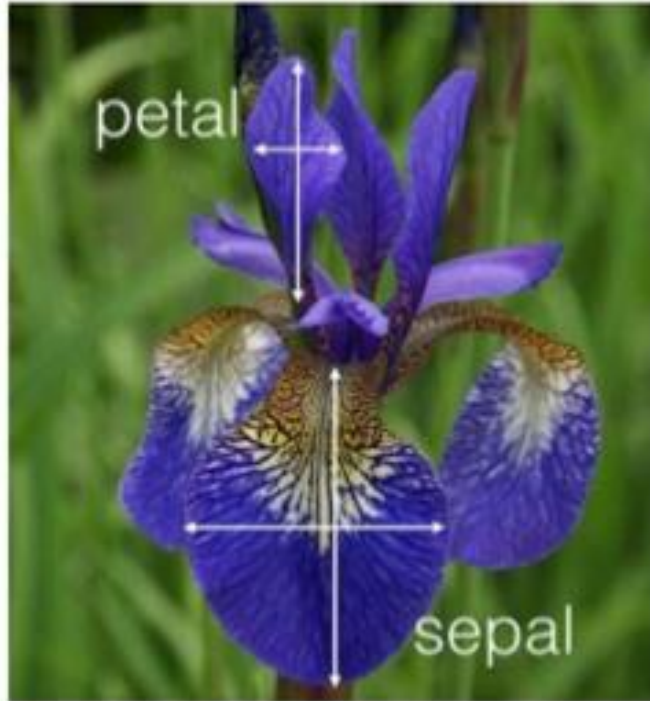


Tipos de Aprendizado

- O **Aprendizado Semi Supervisionado** é uma união do aprendizado **Supervisionado** e o **Não Supervisionado**.
- Esse tipo de aprendizado é capaz de aprender a partir de dados supervisionados e não supervisionados.
- Normalmente usado onde temos uma grande quantidade de dados de treino mas apenas alguns são supervisionados.
- Pode ser usado para gerar dados supervisionados a partir de poucos dados com supervisão



Dataset Iris



Training / test data

Features

Labels

Sepal length	Sepal width	Petal length	Petal width	Species
5.1	3.5	1.4	0.2	Iris setosa
4.9	3.0	1.4	0.2	Iris setosa
7.0	3.2	4.7	1.4	Iris versicolor
6.4	3.2	4.5	1.5	Iris versicolor
6.3	3.3	6.0	2.5	Iris virginica
5.8	3.3	6.0	2.5	Iris virginica

Dataset Iris

IRIS dataset



Iris Versicolor



Iris Virginica



Iris Setosa

Hands on!