

Machine Learning LAB

#1



Abel Eduardo Mondlane

Himani Abel?

- Desenvolvedor de software
- Amante de dados

Do que vamos falar?

Introdução ao Machine Learning

Definição de aprendizado de máquina

Tipos de aprendizado

- Supervisionado vs. Não supervisionado
- Aprendizado por reforço

Algoritmos de Machine Learning

- Visão geral dos algoritmos comuns
- Aplicações do Aprendizado de máquina

Do que vamos falar?

Ambiente de Desenvolvimento

- Introdução ao Google colab
- Introdução ao Python para Machine Learning
- Apresentação das principais bibliotecas (NumPy, Pandas, Scikit-Learn)

O que é ML?

Alguns definem assim:

O aprendizado de máquina (ML) é um ramo da inteligência artificial (IA) e da ciência da computação que se concentra no uso de **dados** e algoritmos para permitir que a IA **imite** a maneira como os **humanos** aprendem, **melhorando gradualmente sua precisão** (*What Is Machine Learning (ML)?, n.d.*).

O que é ML?

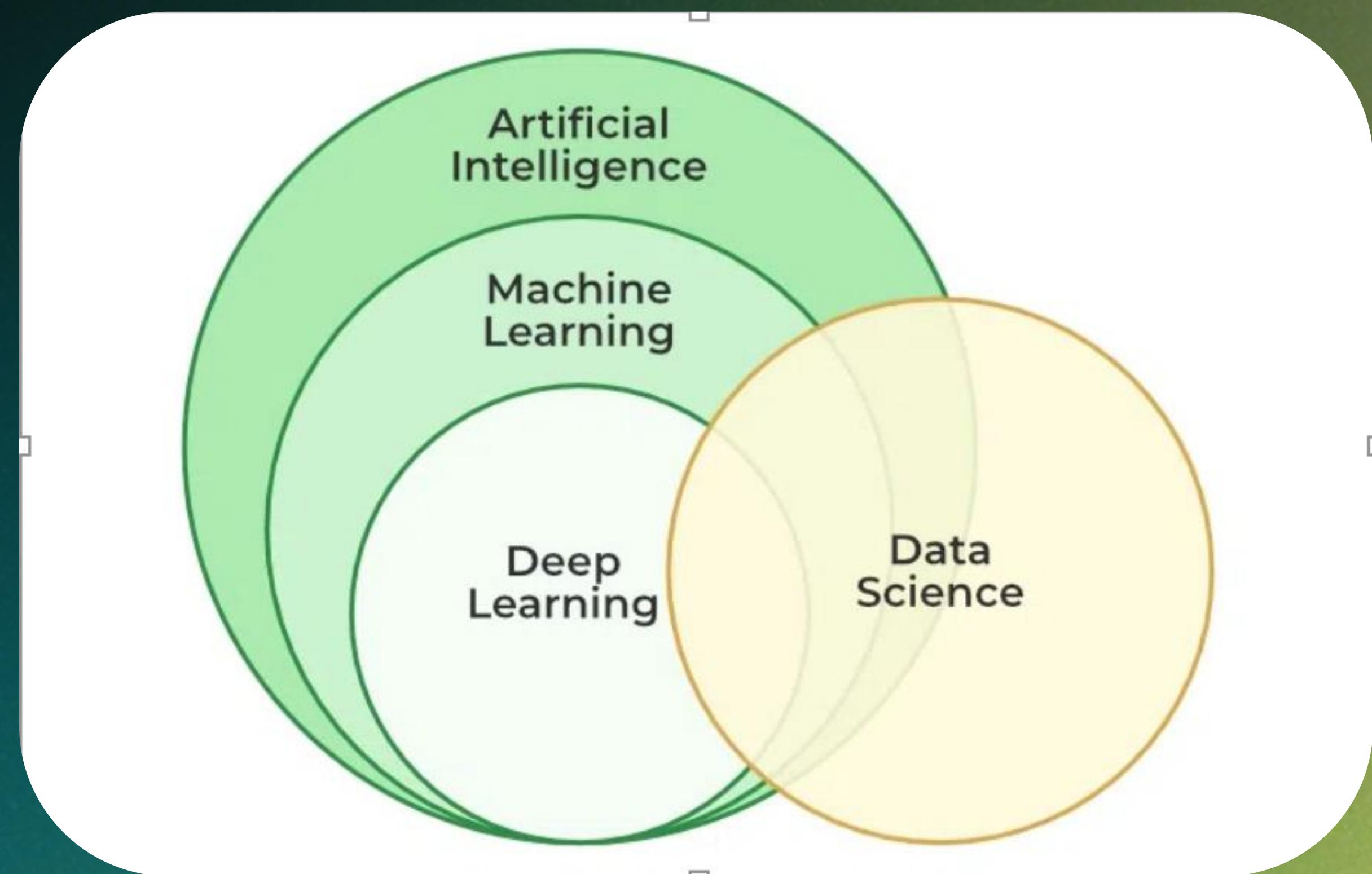
E outros:

O aprendizado de máquina é um subcampo da inteligência artificial, amplamente definido como a capacidade de uma máquina **imitar** o comportamento **humano** inteligente (Brown, 2021).

O que é ML?

E outros ainda...

O que é ML?

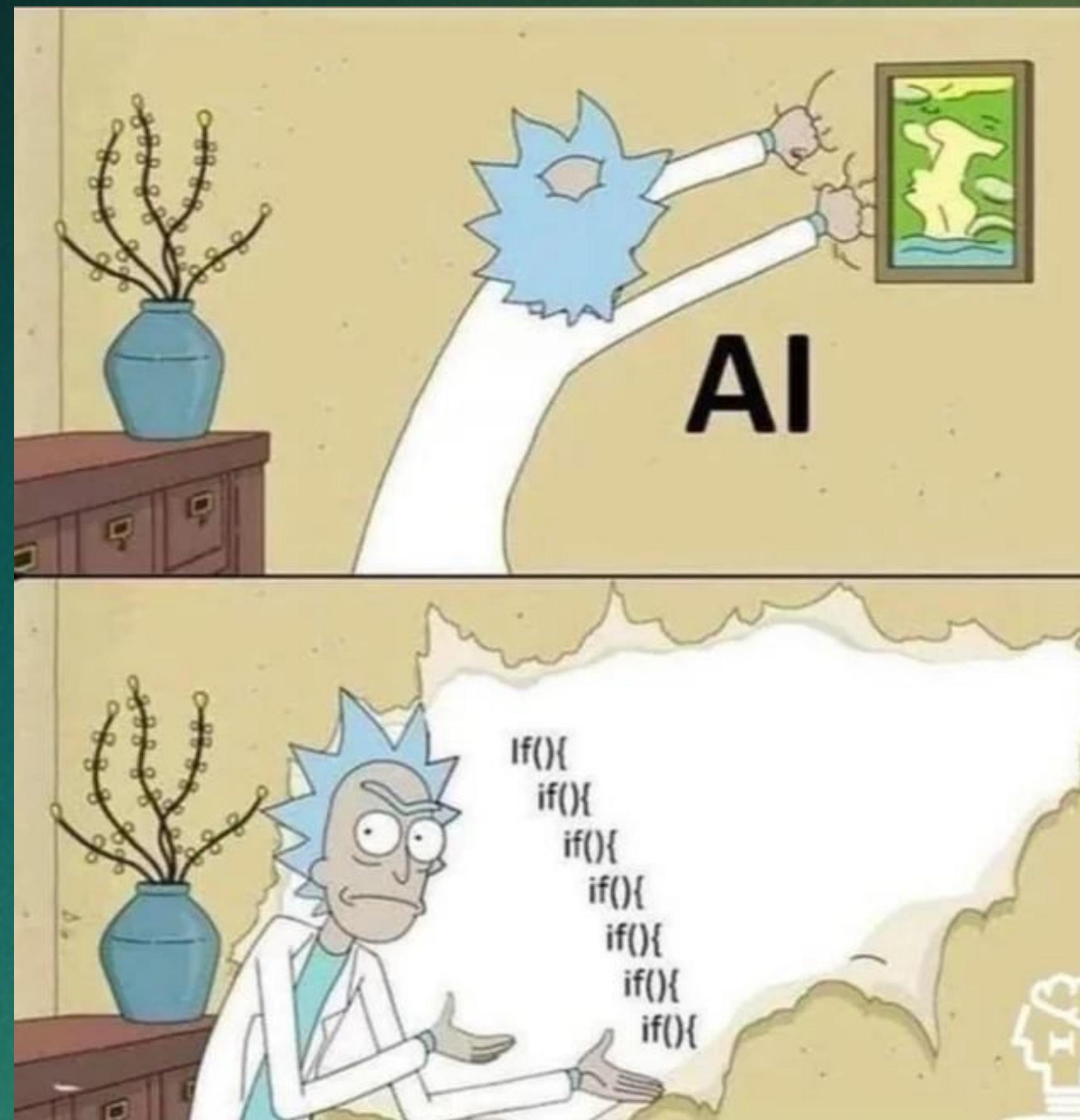


Por que importa?

- Chatbots;
- Predição;
- Tradução;
- Sugestões (de compras, música, vídeos, etc);
- Feeds numa rede social;
- etc.

E...

- Não é explicitamente programado como um software tradicional.



E...

Em alguns casos, escrever um programa para a máquina seguir é demorado ou impossível, como treinar um computador para reconhecer imagens de pessoas diferentes. Embora os humanos possam realizar essa tarefa facilmente, é difícil dizer a um computador como fazê-la.

Tipos de Aprendizado de Máquina

Supervisionado

Sigamos a seguinte analogia:

Na escola, costumamos aprender com a supervisão de um professor.

Nesta abordagem o professor dá a sua matéria e no teste costuma trazer problemas relacionados com o que aprendemos.

Tipos de Aprendizado de Máquina

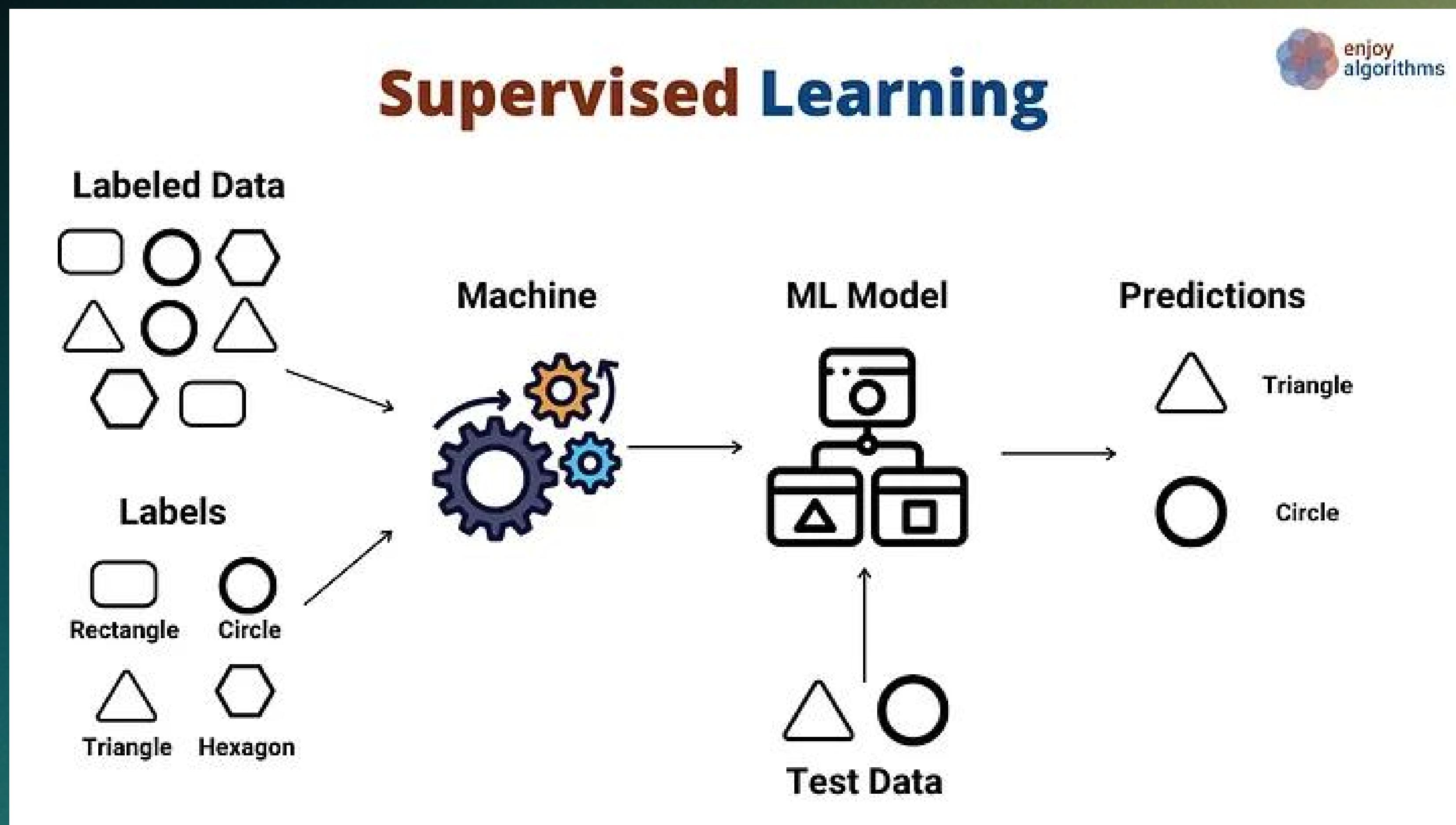
Supervisionado

Não é o tipo de aprendizado supervisionado (das máquinas).

Os computadores são treinados usando exemplos de entrada e saída correspondente. Por exemplo, podemos ter um ficheiro *.csv com várias linhas de dados. Então, cada linha deve conter característica(s) (**features/inputs**) e resultado(s) dessas características (**outputs**).

Tipos de Aprendizado de Máquina

Supervisionado



Tipos de Aprendizado de Máquina

Supervisionado

X1	X2	X3	X4	X5	X6	X7	X8	Y1	Y2	Y3	
0.46165	0.84804	0.13372	0.76001	0.82375	0.41648	0.96926		0	0.53058	0.09042	0.81704
0.60538	0.04752	0.66062	0.98475	0.88622	0.5501	0.41479		1	0.04712	0.85647	0.6926
0.57061	0.16619	0.30468	0.32552	0.35952	0.94022	0.09967		2	0.39981	0.52221	0.85384
0.5731	0.48604	0.16573	0.21886	0.28557	0.53128	0.08253		1	0.90771	0.35441	0.79612
0.20818	0.92313	0.14697	0.14947	0.27753	0.7706	0.51881		2	0.29553	0.17224	0.15541
0.16467	0.25994	0.49221	0.39935	0.61126	0.69532	0.88705		1	0.8814	0.17062	0.19769
0.08073	0.66004	0.39745	0.84362	0.05939	0.99741	0.92084		0	0.24049	0.93115	0.48741
0.15498	0.14142	0.31854	0.32679	0.12081	0.33897	0.36825		0	0.53976	0.49696	0.50545
0.25606	0.63361	0.15927	0.90067	0.43268	0.846	0.26415		0	0.296	0.46516	0.03648
0.00753	0.69324	0.50446	0.24107	0.83918	0.60841	0.69189		2	0.31442	0.51034	0.92371



The diagram illustrates a simple neural network architecture. At the bottom, there is a horizontal line with two curly braces above it. The left brace spans from the first column to the eighth column of the table, labeled "Features". The right brace spans from the ninth column to the eleventh column, labeled "Output". Above the "Features" label, the input layer consists of the columns X1 through X8. Above the "Output" label, the output layer consists of the columns Y1, Y2, and Y3.

Técnicas de aprendizado supervisionado

As duas técnicas de aprendizado supervisionado são classificação e regressão.

Classificação

Os modelos de classificação classificam nossos resultados em determinadas **categorias**. Se o número de categorias for apenas duas, então é especialmente chamada de **classificação binária**. Para maior número de categorias é chamada de **classificação multiclasse**.

Técnicas de aprendizado supervisionado

Classificação

Alguns exemplos de classificação:

- Quais sensores de calor estão a falhar/danificados
- Quais mensagens são spam.

Técnicas de aprendizado supervisionado

Regressão

Os modelos de regressão servem para rotular resultados com valores contínuos.

Alguns exemplos de regressão:

- Prever a temperatura do dia de amanhã;
- Prever quanto tempo a viagem vai durar.

Tipos de Aprendizado de Máquina

Não Supervisionado

O Aprendizado não supervisionado é um método de aprendizado de máquina onde, em contraste com o aprendizado supervisionado, os algoritmos aprendem padrões exclusivamente a partir de **dados não rotulados**.

Tipos de Aprendizado de Máquina

Não Supervisionado

Dentro dessa abordagem, um modelo de aprendizado de máquina tenta encontrar semelhanças, diferenças, padrões e estrutura nos dados por si só. **Nenhuma intervenção humana prévia é necessária** (Unsupervised Learning, n.d.).

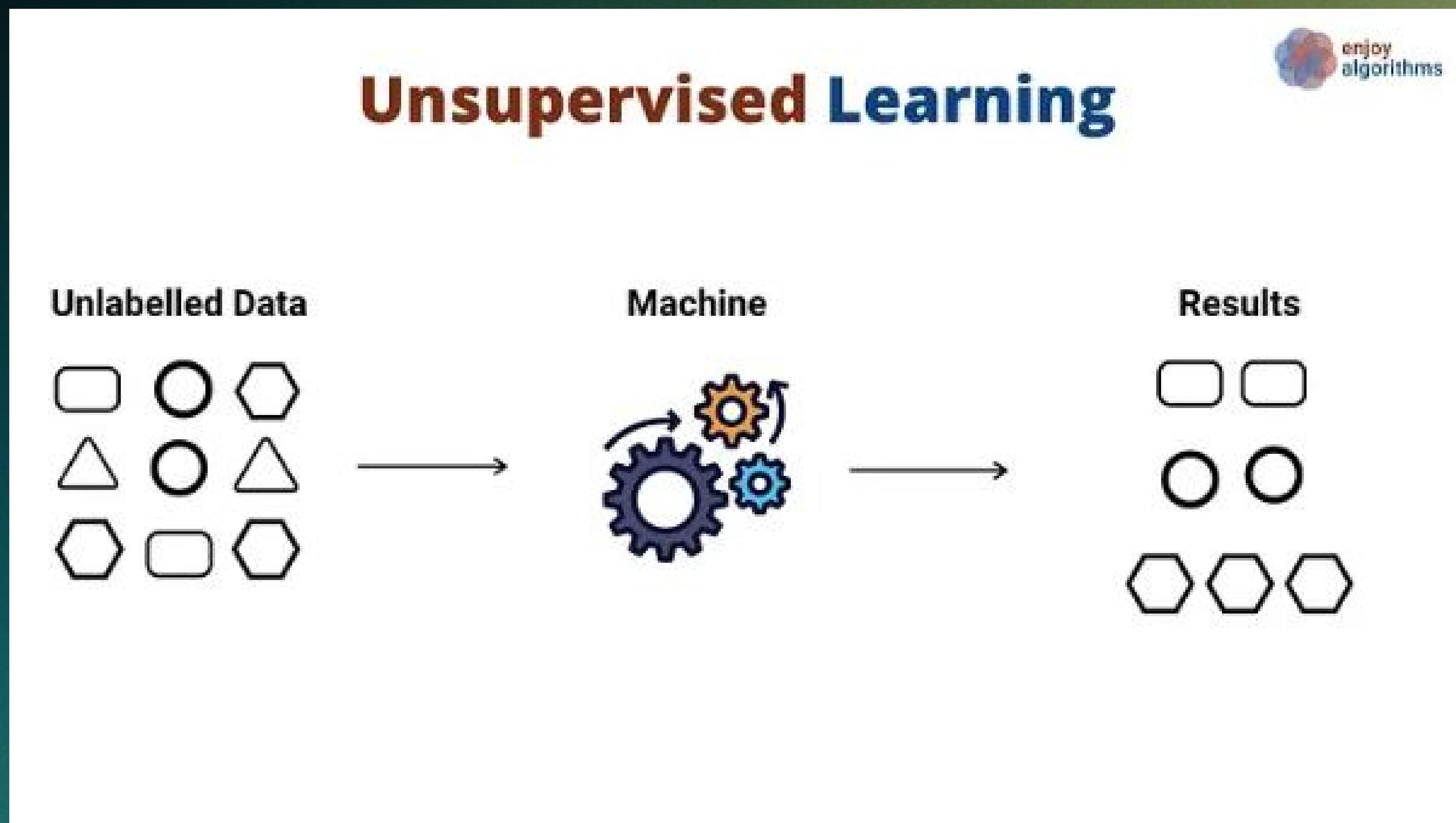
Tipos de Aprendizado de Máquina

Não Supervisionado

Dentro dessa abordagem, um modelo de aprendizado de máquina tenta encontrar semelhanças, diferenças, padrões e estrutura nos dados por si só. **Nenhuma intervenção humana prévia é necessária** (Unsupervised Learning, n.d.).

Tipos de Aprendizado de Máquina

Não Supervisionado



Técnicas de aprendizado Não supervisionado

As duas técnicas de aprendizado não supervisionado mais conhecidas são clustering e association.

* nesta sessão falaremos apenas de clustering.

Técnicas de aprendizado Não supervisionado

Clustering

Ao contrário da classificação e da regressão, o agrupamento não envolve a atribuição de rótulos predefinidos ou a previsão de valores. Em vez disso, o clustering agrupa pontos de dados com base na semelhança de seus recursos. O objetivo é identificar padrões e relacionamentos inerentes aos dados. O clustering é particularmente útil quando temos dados não rotulados e queremos descobrir estruturas ocultas (Bathula, 2023).

Técnicas de aprendizado Não supervisionado

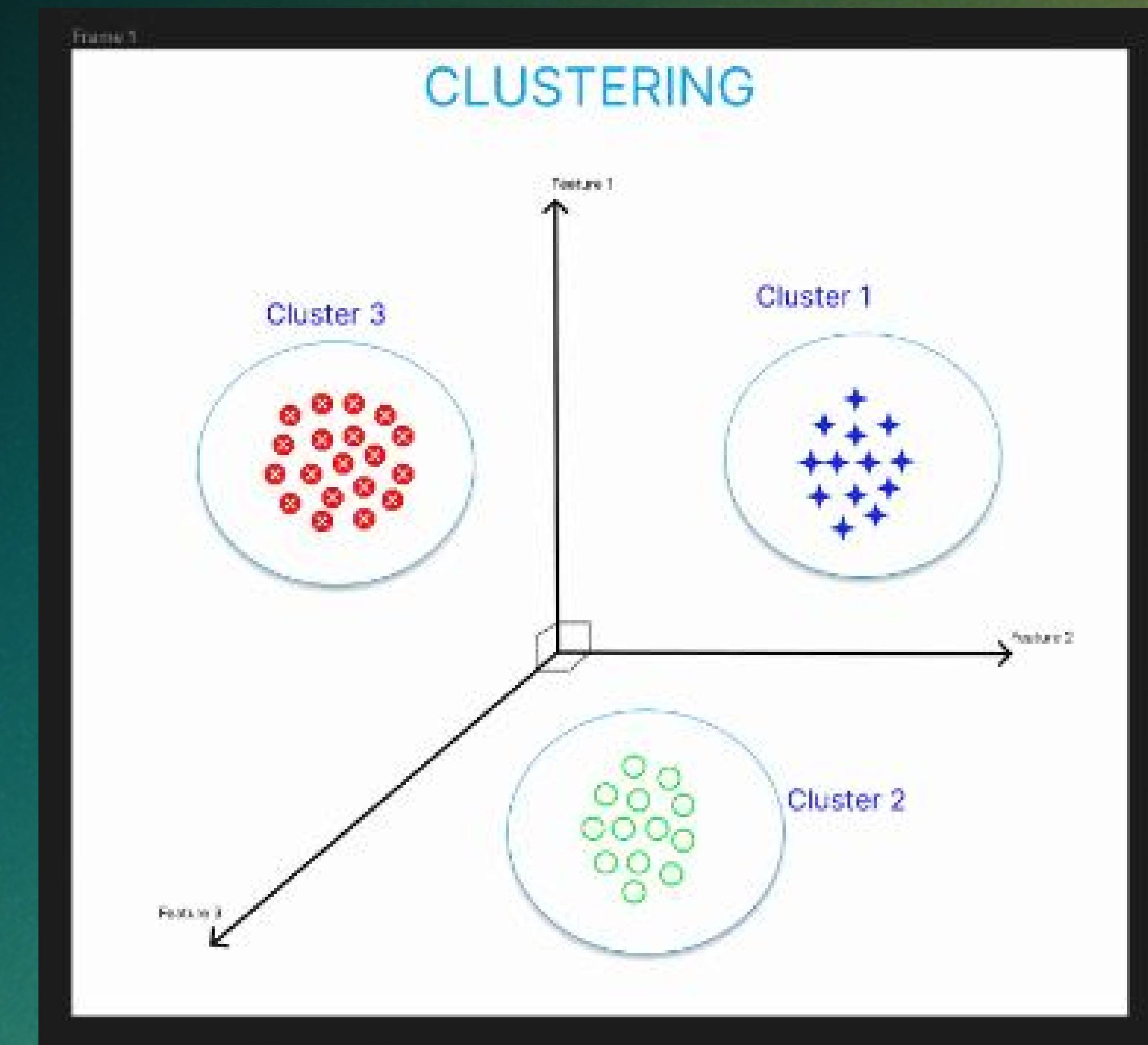
Clustering

Exemplos de clustering:

- De acordo com as características do programadores, definir seus respectivos grupos (frontend,backend,...)
- Agrupar clientes com características/interesses semelhantes.

Técnicas de aprendizado Não supervisionado

Clustering



Algoritmos de Machine Learning

Decision Tree

O algoritmo de árvore de decisão (decision tree) é um método popular de aprendizado supervisionado em aprendizado de máquina. É um tipo de modelo que utiliza uma estrutura semelhante a uma **árvore** para fazer previsões com base em recursos de entrada. Em uma árvore de decisão, cada nó representa um recurso/atributo e cada ramo representa um valor possível desse recurso (Desai, 2023).

Algoritmos de Machine Learning

Decision Tree

O algoritmo de árvore de decisão (decision tree) é um método popular de aprendizado supervisionado em aprendizado de máquina. É um tipo de modelo que utiliza uma estrutura semelhante a uma **árvore** para fazer previsões com base em recursos de entrada. Em uma árvore de decisão, cada nó representa um recurso/atributo e cada ramo representa um valor possível desse recurso (Desai, 2023).

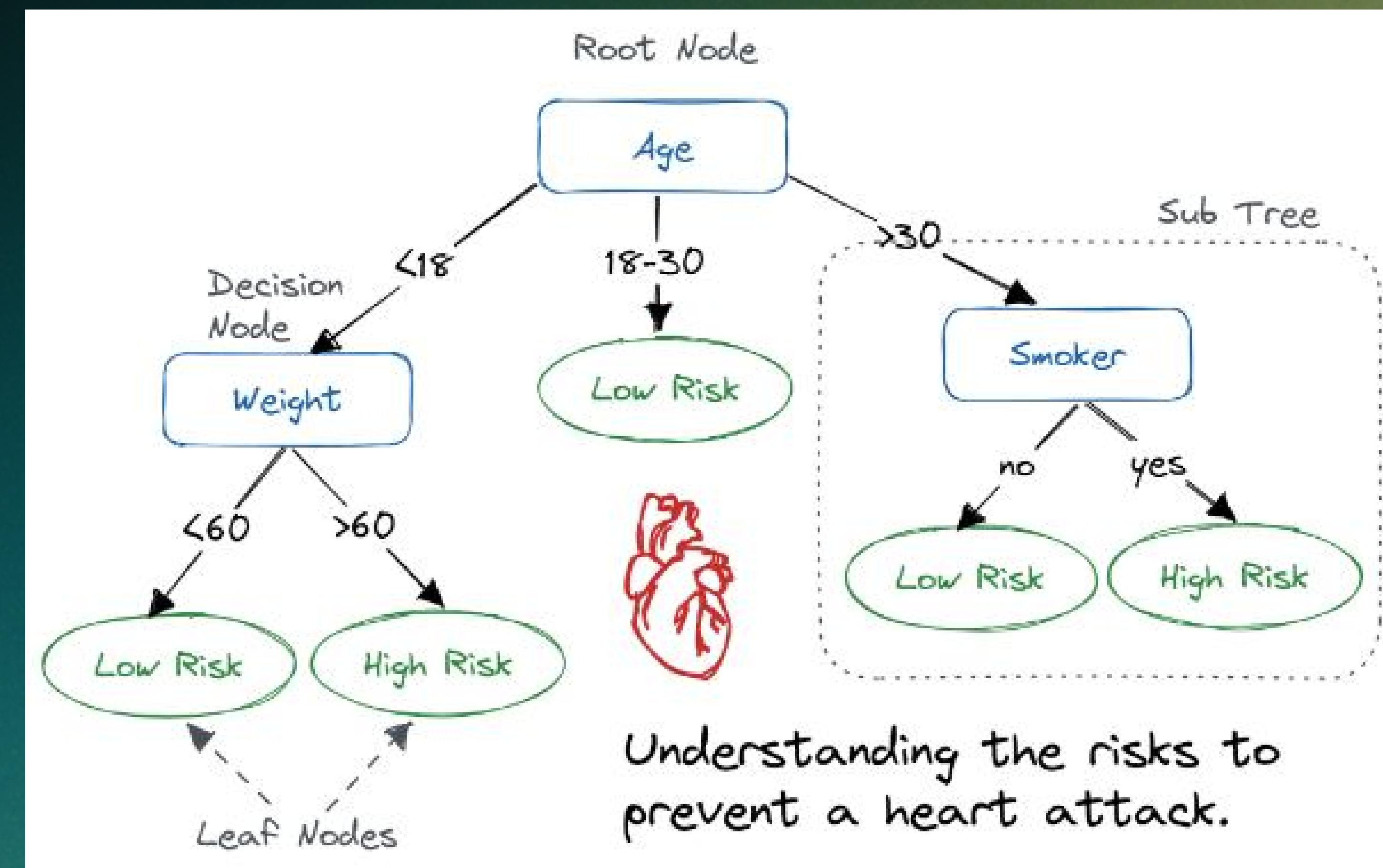
Algoritmos de Machine Learning

Decision Tree

O algoritmo de árvore de decisão (decision tree) é um método popular de aprendizado supervisionado em aprendizado de máquina. É um tipo de modelo que utiliza uma estrutura semelhante a uma **árvore** para fazer previsões com base em recursos de entrada. Em uma árvore de decisão, cada nó representa um recurso/atributo e cada ramo representa um valor possível desse recurso (Desai, 2023).

Algoritmos de Machine Learning

Decision Tree



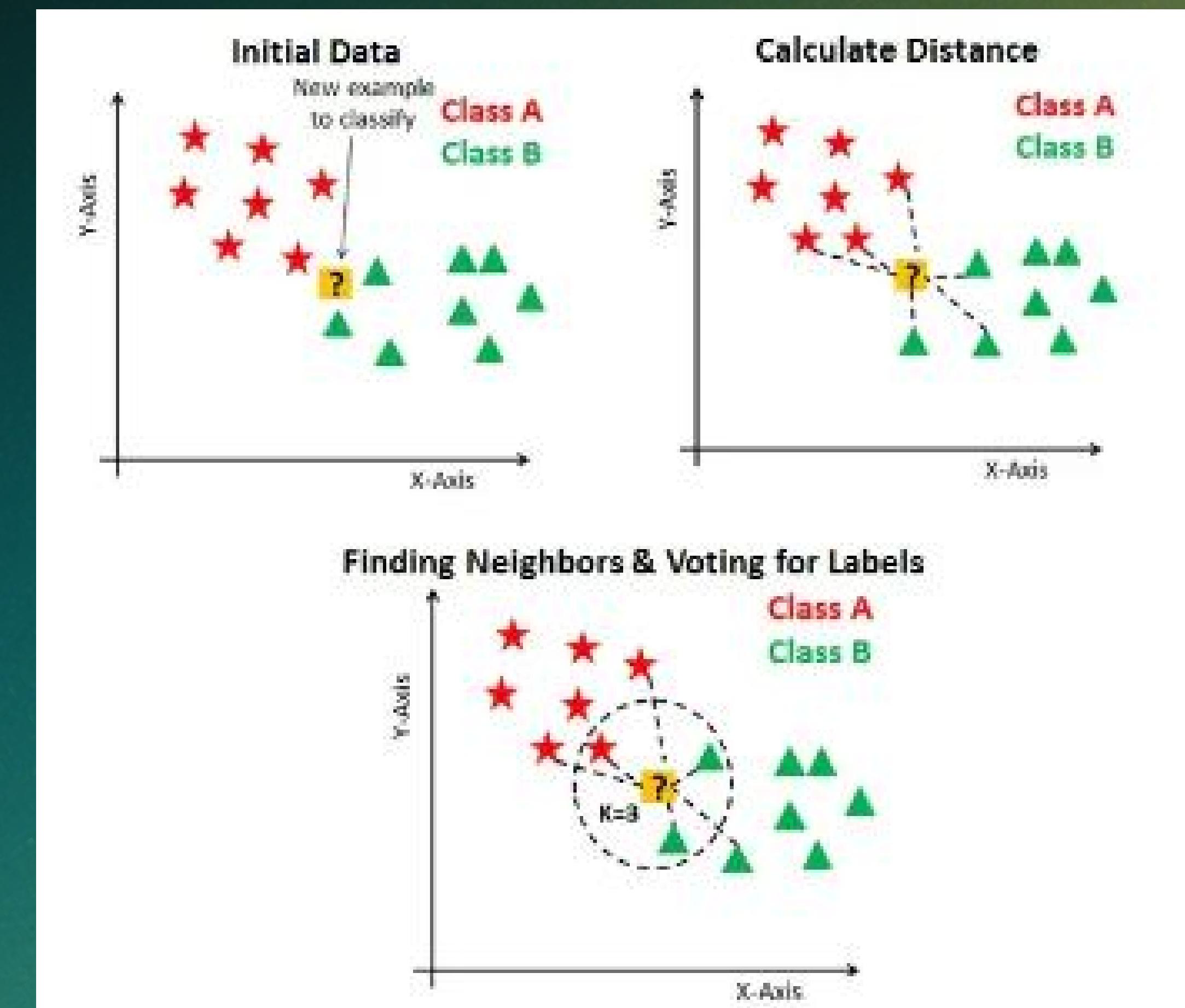
Algoritmos de Machine Learning

K-Nearest Neighbors

O algoritmo KNN é baseado no princípio de que objetos ou pontos de dados próximos uns dos outros no espaço de recursos provavelmente pertencem à mesma classe ou têm saídas semelhantes. O algoritmo KNN atribui um novo ponto de dados à classe que é mais comum entre seus k vizinhos mais próximos. O valor de k é um hiperparâmetro que precisa ser especificado antes do treinamento do modelo (Desai, 2023).

Algoritmos de Machine Learning

K-Nearest Neighbors



Algoritmos de Machine Learning

Linear Regression

A regressão linear é usada para identificar a relação entre uma variável dependente e uma ou mais variáveis independentes e normalmente é aproveitada para fazer previsões sobre resultados futuros. Quando há apenas uma variável independente e uma variável dependente, isso é conhecido como regressão linear simples (What Is a Machine Learning Algorithm?, n.d.).