

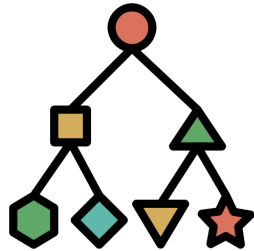
# Aula de Big Data



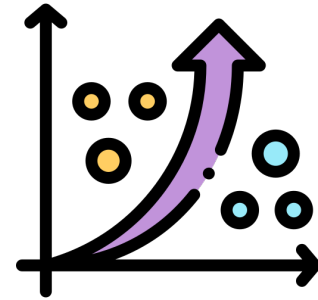
*Alo-Você*  
Name = **Leandro de Souza**  
Instagram = **@prof.leandrodesouza**

AULA07

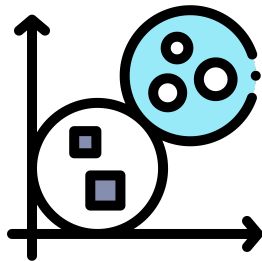
# TAREFAS



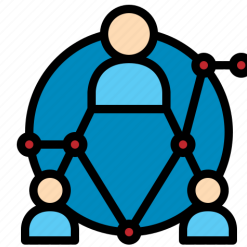
**CLASSIFICAÇÃO**



**REGRESSÃO**

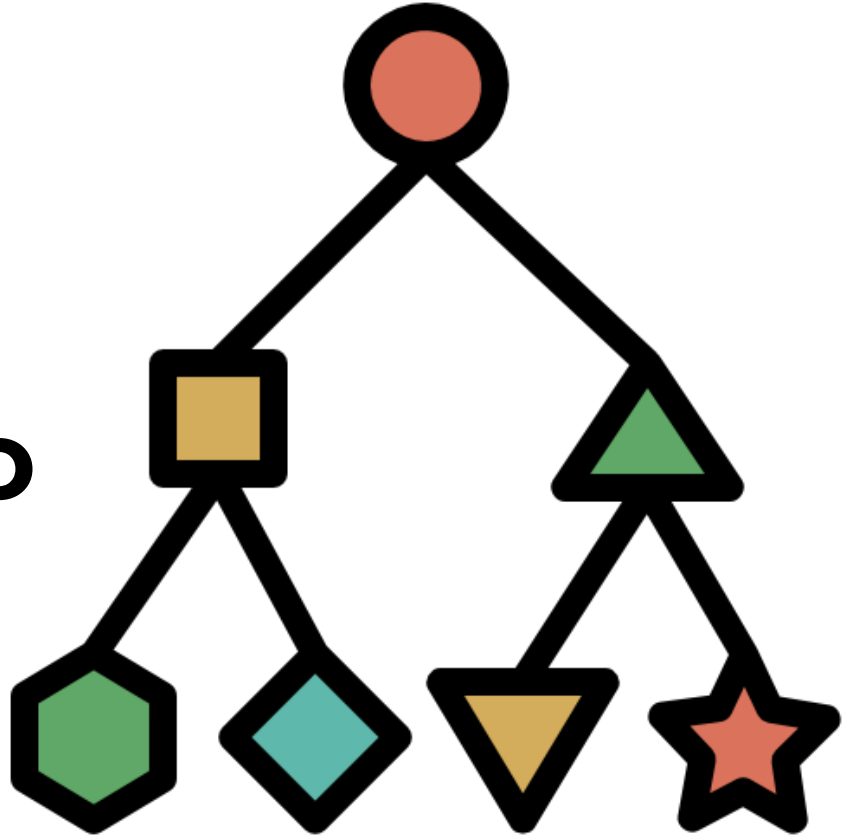


**AGRUPAMENTOS**



**REGRAS DE ASSOCIAÇÃO**

# CLASSIFICAÇÃO



# CLASSIFICAÇÃO

Na classificação queremos: **descobrir ou descrever a classe de um fato.**

A **classe** em uma relação está representada em um **atributo especial**, posicionado como **último atributo da relação**.

# CLASSIFICAÇÃO

É a tarefa mais comum em mineração de dados

Tem como objetivo utilizar atributos de um objeto para determinar a qual classe ele pertence

Tenta prever à qual classe pertence uma nova instância

# CLASSIFICAÇÃO

ID ↕	Comprimento sépala ↕	Largura sépala ↕	Comprimento pétala ↕	Largura pétala ↕	Especie ↕
1	5.1	3.5	1.4	0.2	<i>I. setosa</i>
2	4.9	3.0	1.4	0.2	<i>I. setosa</i>
3	4.7	3.2	1.3	0.2	<i>I. setosa</i>
4	4.6	3.1	1.5	0.2	<i>I. setosa</i>
5	5.0	3.6	1.4	0.3	<i>I. setosa</i>
6	5.4	3.9	1.7	0.4	<i>I. setosa</i>
7	4.6	3.4	1.4	0.3	<i>I. setosa</i>
8	5.0	3.4	1.5	0.2	<i>I. setosa</i>
9	4.4	2.9	1.4	0.2	<i>I. setosa</i>
10	4.9	3.1	1.5	0.1	<i>I. setosa</i>
11	5.4	3.7	1.5	0.2	<i>I. setosa</i>
12	4.8	3.4	1.6	0.2	<i>I. setosa</i>
13	4.8	3.0	1.4	0.1	<i>I. setosa</i>
14	4.3	3.0	1.1	0.1	<i>I. setosa</i>
15	5.8	4.0	1.2	0.2	<i>I. setosa</i>
16	5.7	4.4	1.5	0.4	<i>I. setosa</i>
17	5.4	3.9	1.3	0.4	<i>I. setosa</i>
18	5.1	3.5	1.4	0.3	<i>I. setosa</i>
19	5.7	3.8	1.7	0.3	<i>I. setosa</i>
20	5.1	3.8	1.5	0.3	<i>I. setosa</i>

Dados de Iris de Fisher

# CLASSIFICAÇÃO

## APRENDIZADO SUPERVISIONADO

São exemplos de algoritmos de classificação: árvores de decisão, classificação Bayesiana, classificação baseada em regras, máquinas de vetores suporte (*support vector machines*) e redes neurais.

# MATRIZ DE CONFUSÃO

A Matriz de Confusão é calculada pela função de pesquisa Classificação.

Ela exibe a **distribuição dos registros em termos de suas classes atuais e de suas classes previstas.**

Isso indica a qualidade do modelo atual.

Um modelo pode conter duas ou mais classes previstas.

Valor real \ Valor previsto	Positivo	Negativo
Positivo	Verdadeiro Positivo (TP)	Falso Positivo (FP)
Negativo	Falso Negativo (FN)	Verdadeiro negativo (TN)



# MATRIZ DE CONFUSÃO



Valor previsto \ Valor real	Positivo	Negativo
Positivo	Verdadeiro Positivo (TP)	Falso Positivo (FP)
Negativo	Falso Negativo (FN)	Verdadeiro negativo (TN)

- **Verdadeiros Positivos (TP)** são observações cujo valor real é positivo e o valor previsto é positivo, isto é, o modelo acertou
- **Verdadeiros Negativos (TN)** são observações cujo valor real é negativo e o valor previsto é negativo, isto é, o modelo acertou

# MATRIZ DE CONFUSÃO

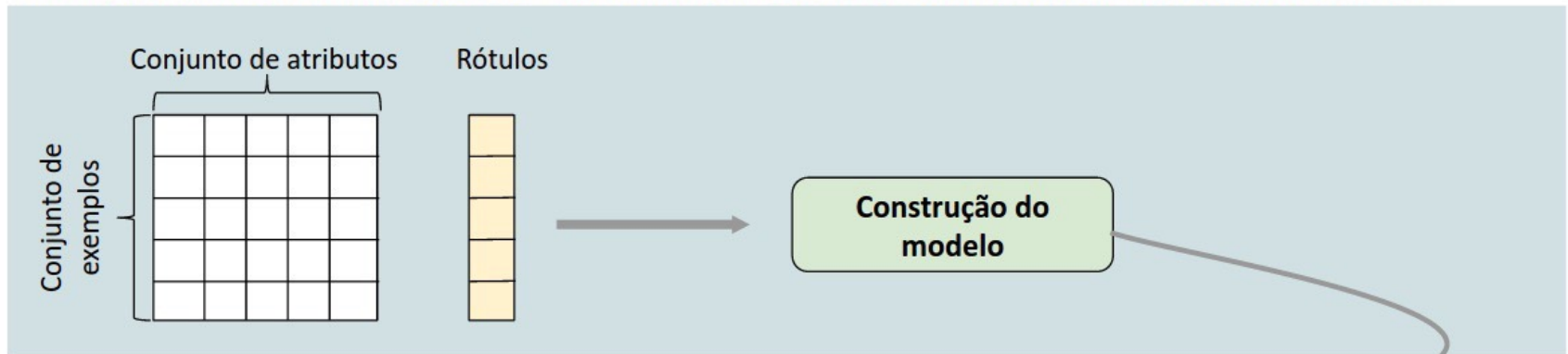
Valor previsto \ Valor real	Positivo	Negativo
Positivo	Verdadeiro Positivo (TP)	Falso Positivo (FP)
Negativo	Falso Negativo (FN)	Verdadeiro negativo (TN)

- **Falsos Positivos (FP)** são casos em que o resultado correto é negativo entretanto o resultado obtido é positivo, isto é, o modelo errou
- **Falsos Negativos (FN)** são casos em que o resultado correto é positivo entretanto o resultado obtido é negativo, isto é, o modelo errou

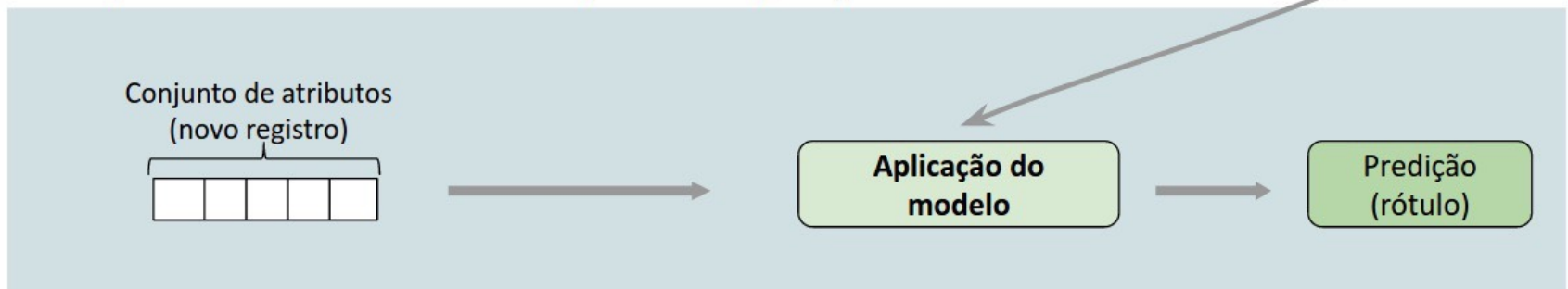
# MATRIZ DE CONFUSÃO

		Predição	
		P	N
Real	p	TP	FN
	n	FP	TN

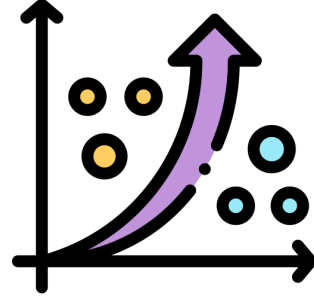
**Fase de treinamento:** um algoritmo processa um conjunto de dados rotulados, gerando um modelo



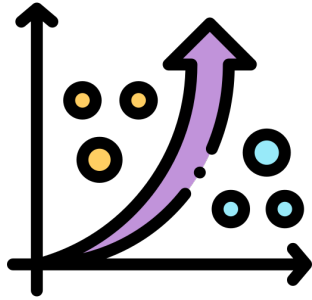
**Fase operacional:** o modelo é utilizado para fazer a predição de dados ainda não rotulados



# REGRESSÃO







# REGRESSÃO

## Objetivo:

Aqui, buscamos prever a classe de forma **NUMÉRICA**

**Exemplos:** Prever a altura de uma pessoa a partir do peso.

# REGRESSÃO

Busca prever um valor numérico contínuo.

A resposta desejada é um valor contínuo, e não um rótulo do tipo "sim/não"



# REGRESSÃO

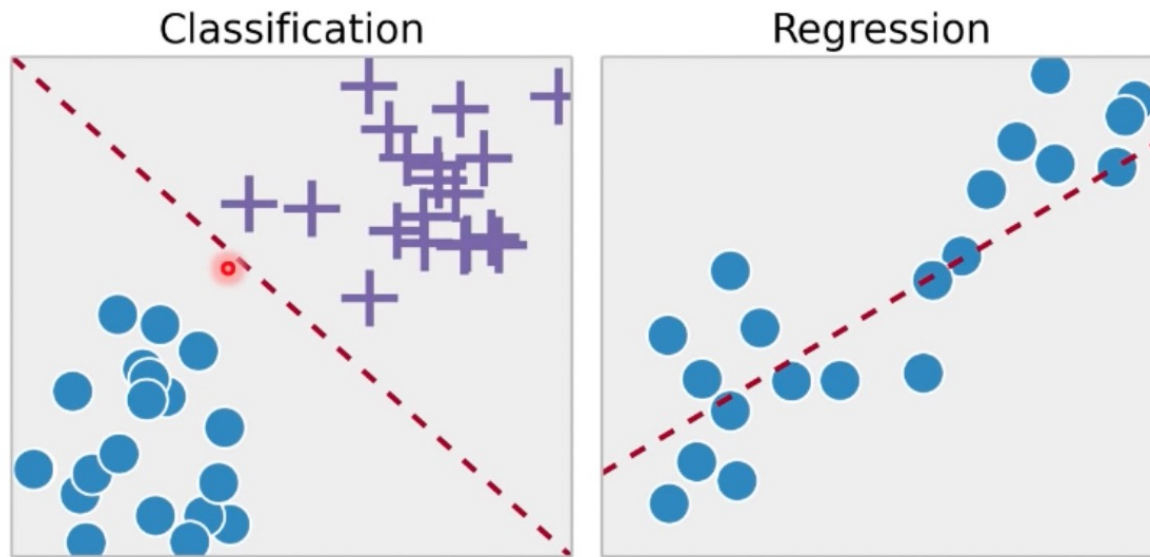
São exemplos de algoritmos de regressão:

- regressão linear simples e múltipla
- regressão não linear simples e múltipla

# REGRESSÃO

Modelagem da relação entre variáveis numéricas.

Variável dependente **y** e variáveis explanatórias **x**



# REGRESSÃO

**X**

**y**

Temperatura, umidade e  
pressão do ar



Velocidade do vento

Gastos com cartão de crédito,  
histórico de pagamentos



Limite de crédito

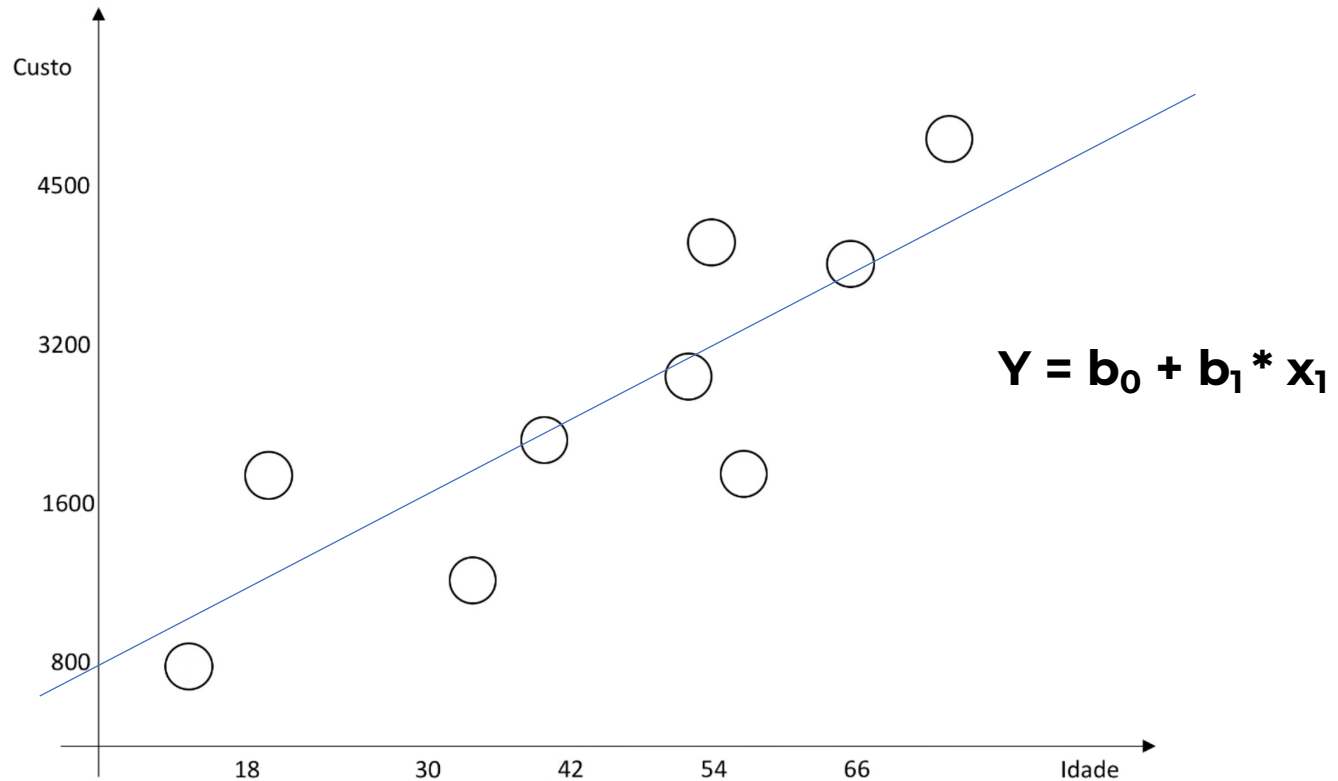
Idade



Custo plano de saúde

# REGRESSÃO LINEAR SIMPLES

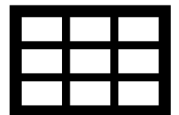
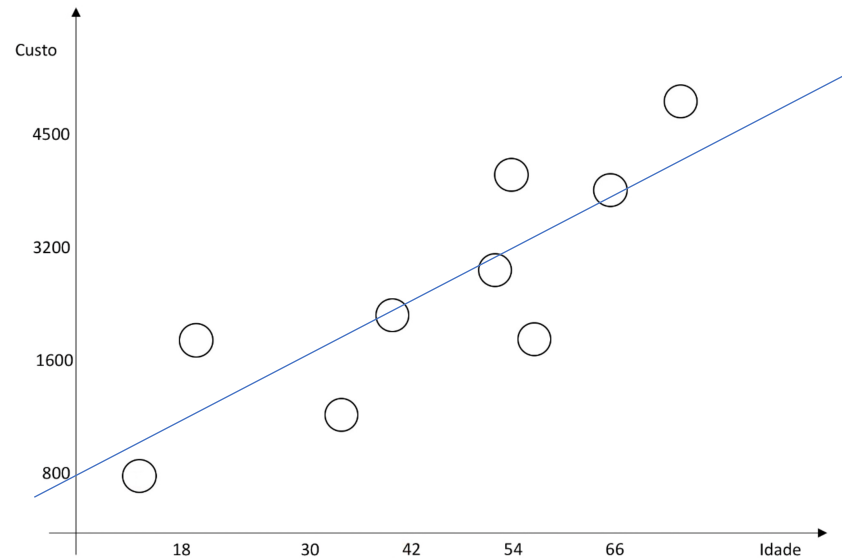
Apenas uma variável explanatória **x**



# REGRESSÃO LINER SIMPLES

$$Y = b_0 + b_1 * x_1$$

Relação linear entre os atributos: quanto maior a idade, maior o custo  **$b_0$**  e  **$b_1$**  definem a localização da linha.



# Qual a diferença entre cada uma dessas tarefas?

Preparando os  
**DADOS**



Qual será a  
temperatura  
amanhã?



Vai fazer frio  
ou calor  
amanhã?

**Regressão** — **Classificação**