

Aprendizado de Máquina Supervisionado



Departamento de
Computação - **UFSCar**



Aprendizado Supervisionado

- Aprende com uma base de dados de treino com as “respostas certas”.

X



Y

**input/features/
Dados**



**output/target/
Previsão**



Departamento de
Computação - **UFSCar**



Aplicações

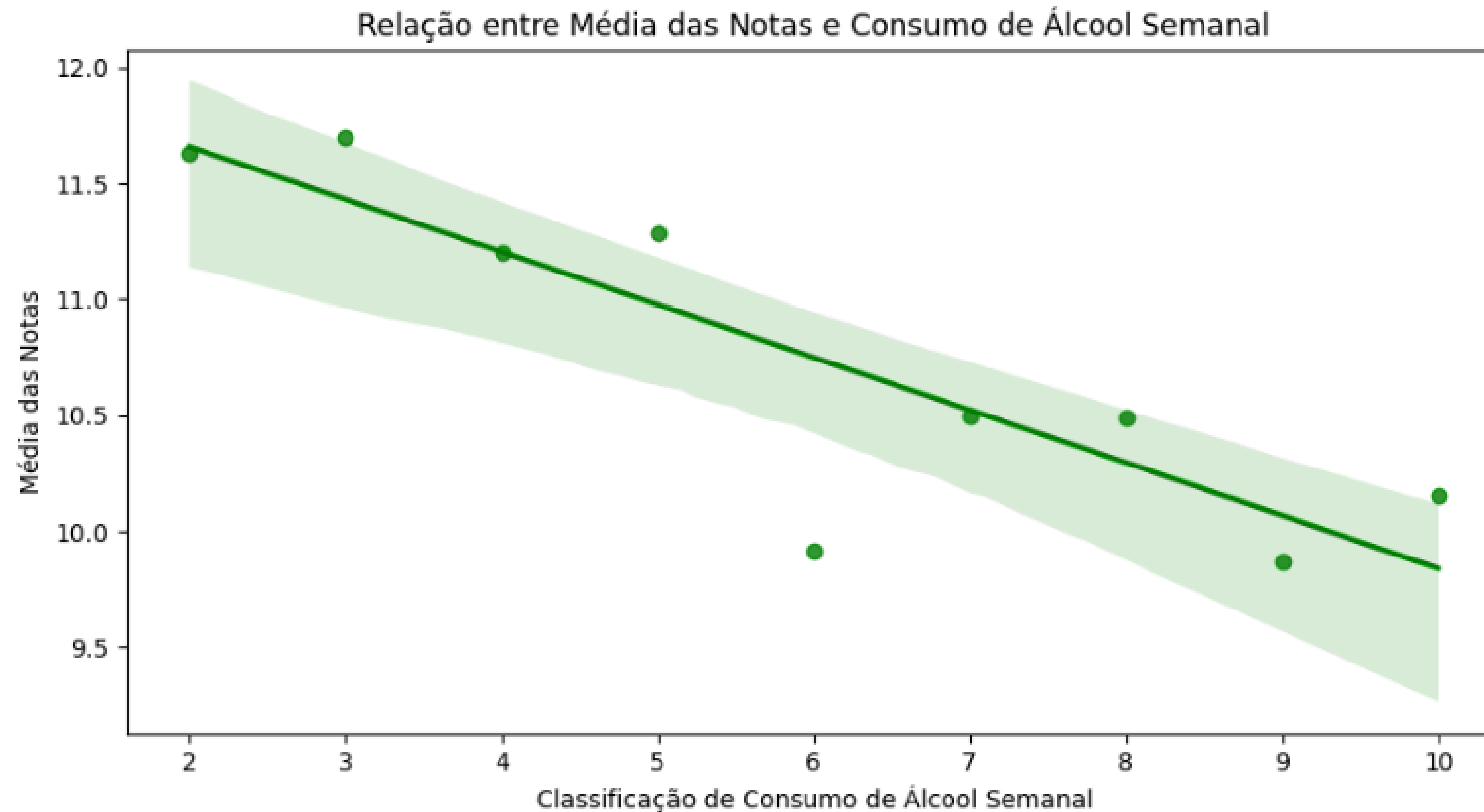
Input(X)	Output(Y)	Aplicação
email	spam ? (1/0)	Filtro de Spam
imagem	cachorro ou gato? (1/0)	Visão Computacional
English	Spanish	Machine Translation
Anúncio	Cliclou ? (1/0)	Venda de produtos online



Departamento de
Computação - **UFSCar**



Regressão



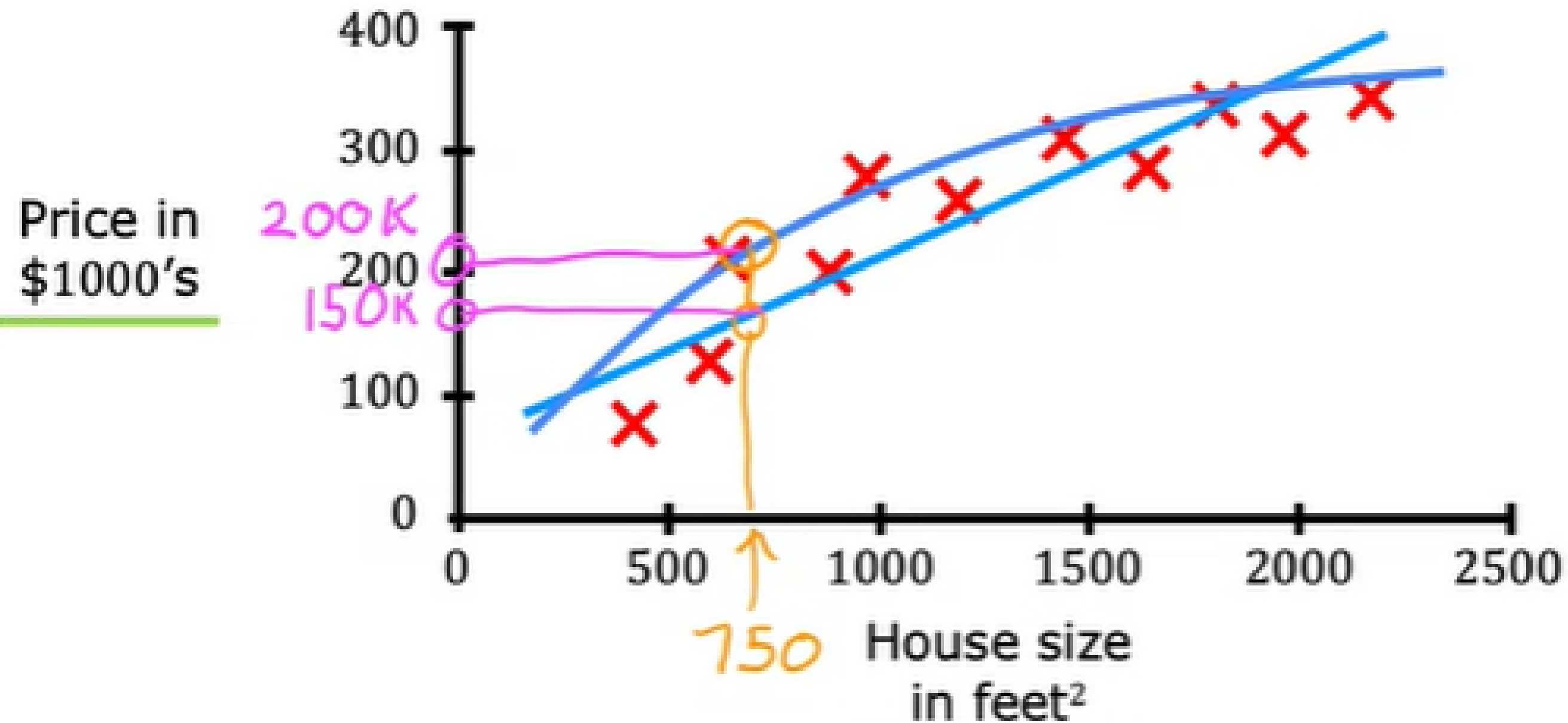
Quando usa uma regressão linear ou uma função mais complexa?



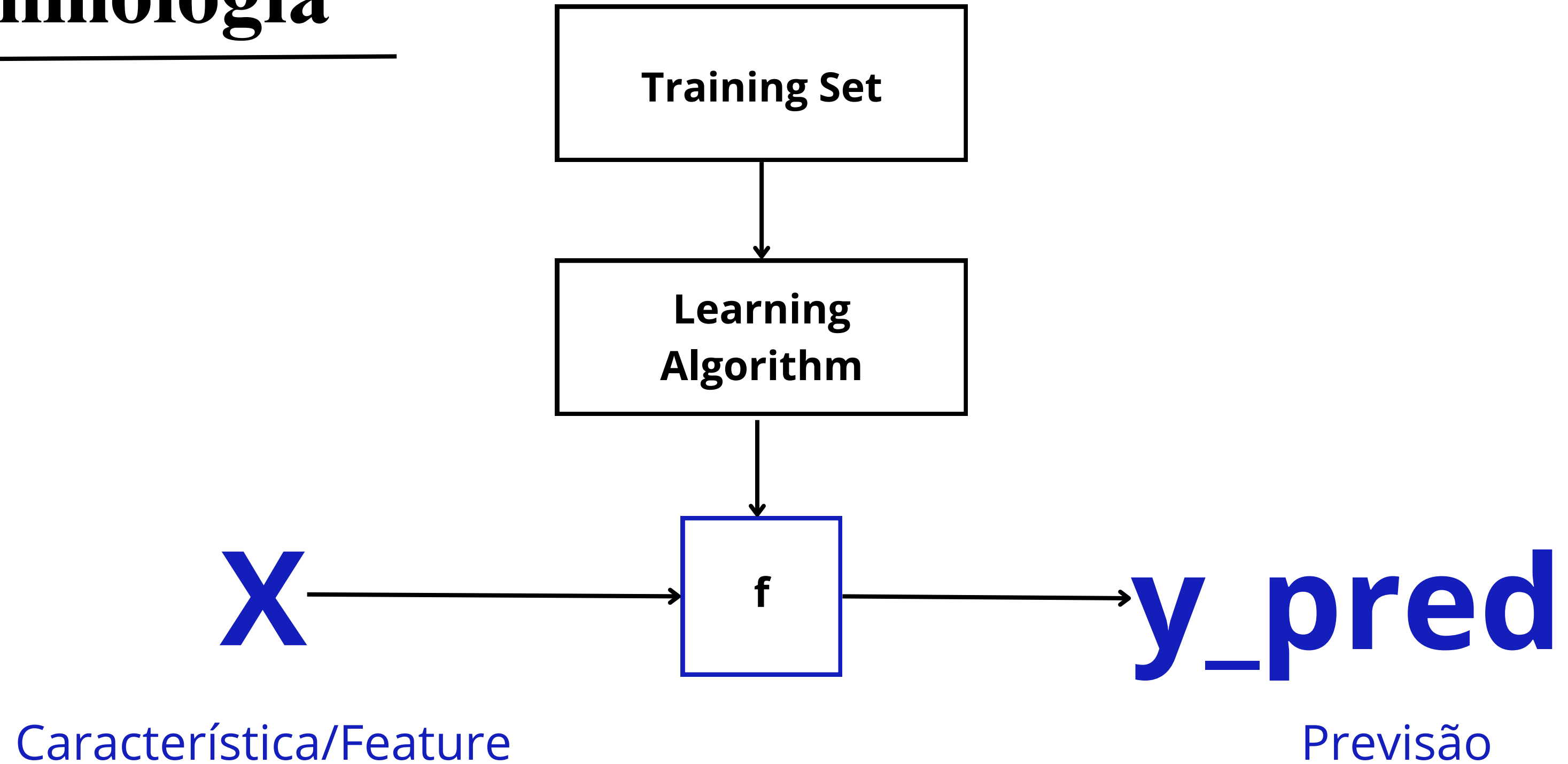
Departamento de
Computação - **UFSCar**



Regression: Housing price prediction



Terminologia



Departamento de
Computação - **UFSCar**



Classificação

- Prevê categorias, ou seja, um número bem limitado de outputs.
- Por exemplo, inferir se um paciente tem um problema do coração baseado em dados como: maior bpm atingindo, nível de colesterol, predisposição genética, peso, idade, sexo...
- Notebook: Modelo que infere se um estudante largou os estudos/está cursando/terminou a faculdade com base em dados demográficos e sociais



Departamento de
Computação - **UFSCar**

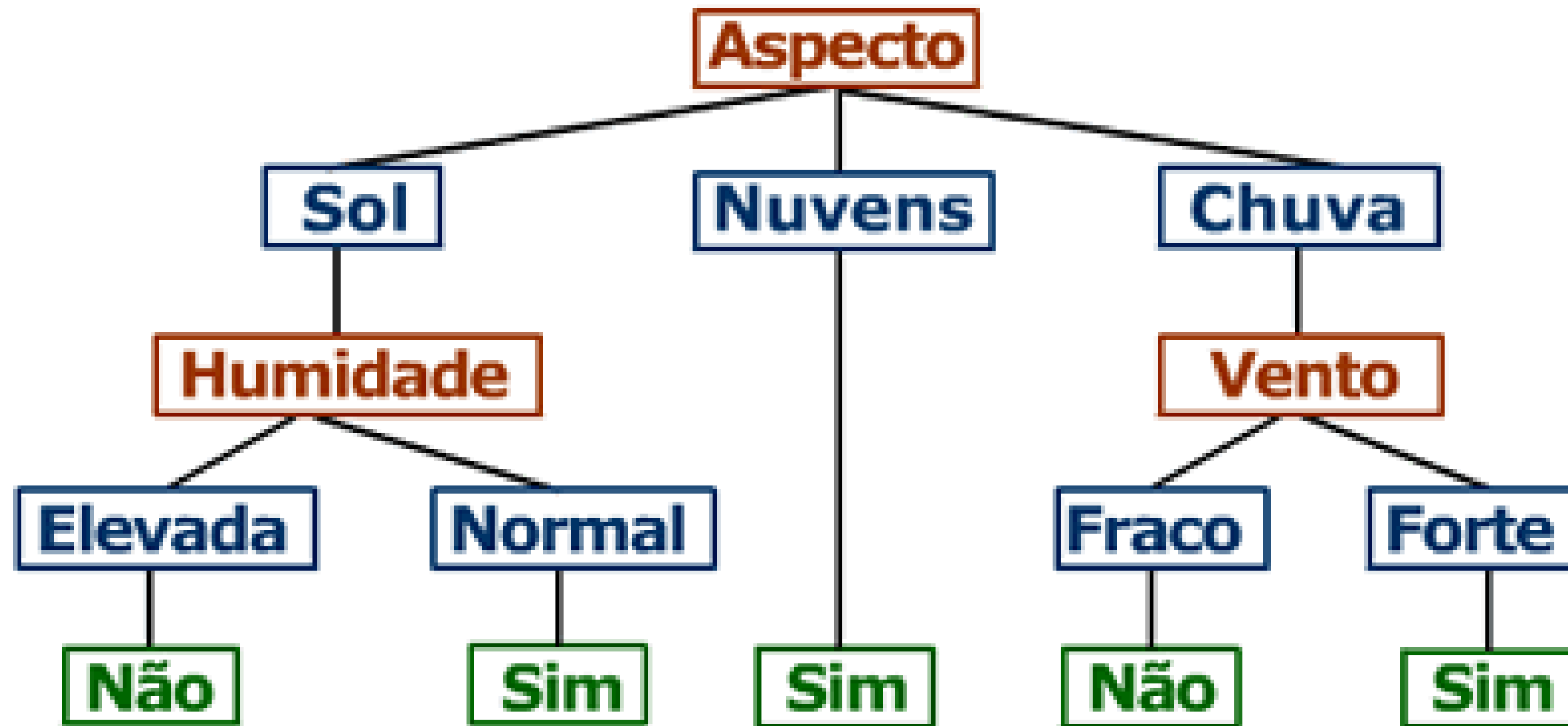


Exemplos de Treino



Dia	Aspecto	Temp.	Humidade	Vento	Jogar Tênis
D1	Sol	Quente	Elevada	Fraco	Não
D2	Sol	Quente	Elevada	Forte	Não
D3	Nuvens	Quente	Elevada	Fraco	Sim
D4	Chuva	Ameno	Elevada	Fraco	Sim
D5	Chuva	Fresco	Normal	Fraco	Sim
D6	Chuva	Fresco	Normal	Forte	Não
D7	Nuvens	Fresco	Normal	Fraco	Sim
D8	Sol	Ameno	Elevada	Fraco	Não
D9	Sol	Fresco	Normal	Fraco	Sim
D10	Chuva	Ameno	Normal	Forte	Sim
D11	Sol	Ameno	Normal	Forte	Sim
D12	Nuvens	Ameno	Elevada	Forte	Sim
D13	Nuvens	Quente	Normal	Fraco	Sim
D14	Chuva	Ameno	Elevada	Forte	Não

Algoritmo ID3

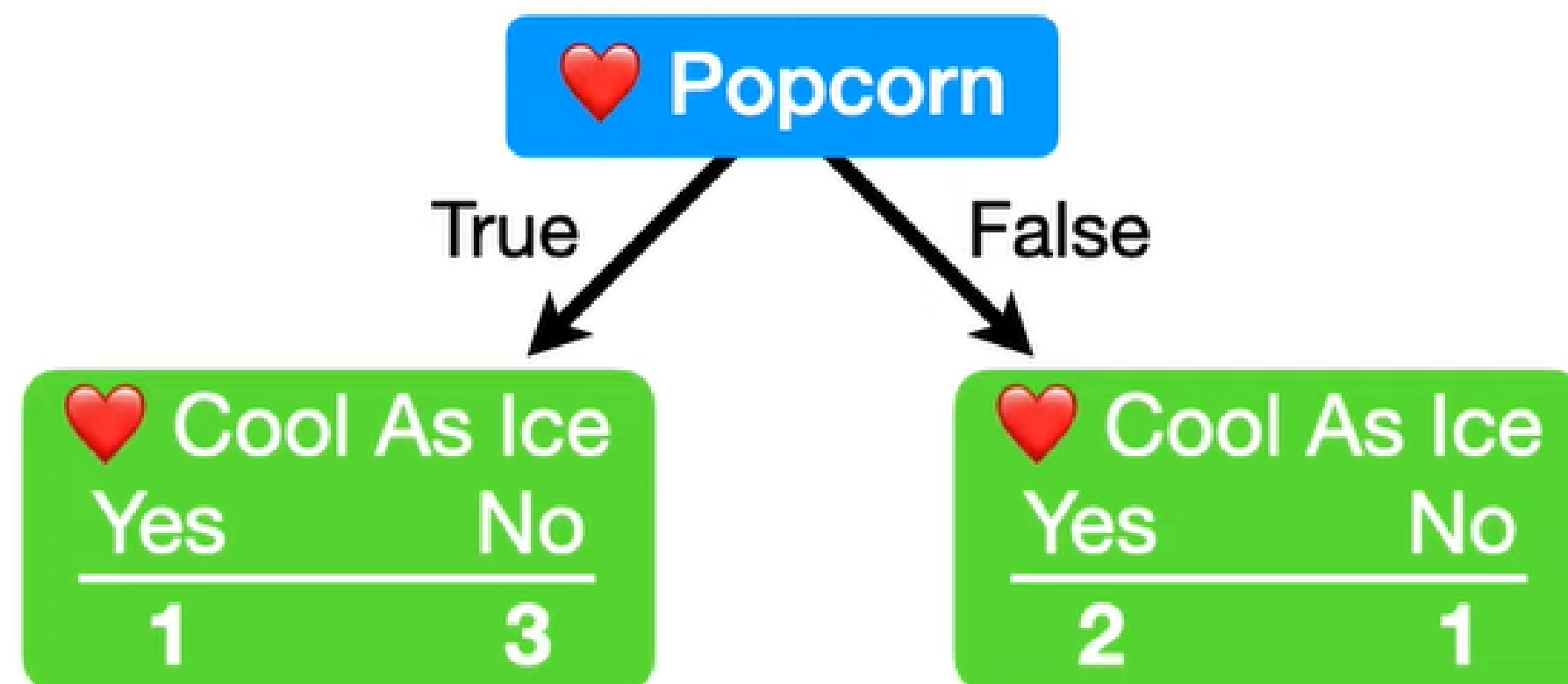
Árvore de Decisão para Jogar Tênis



Impureza de Gini

Loves Popcorn	Loves Soda	Age	Loves Cool As Ice
Yes	Yes	7	No
Yes	No	12	No
No	Yes	18	Yes
No	Yes	35	Yes
Yes	Yes	38	Yes
Yes	No	50	No
No	No	83	No



Gini Impurity for a Leaf = $1 - (\text{the probability of "Yes"})^2 - (\text{the probability of "No"})^2$

$$= 1 - \left(\frac{1}{1+3}\right)^2 - \left(\frac{3}{1+3}\right)^2$$

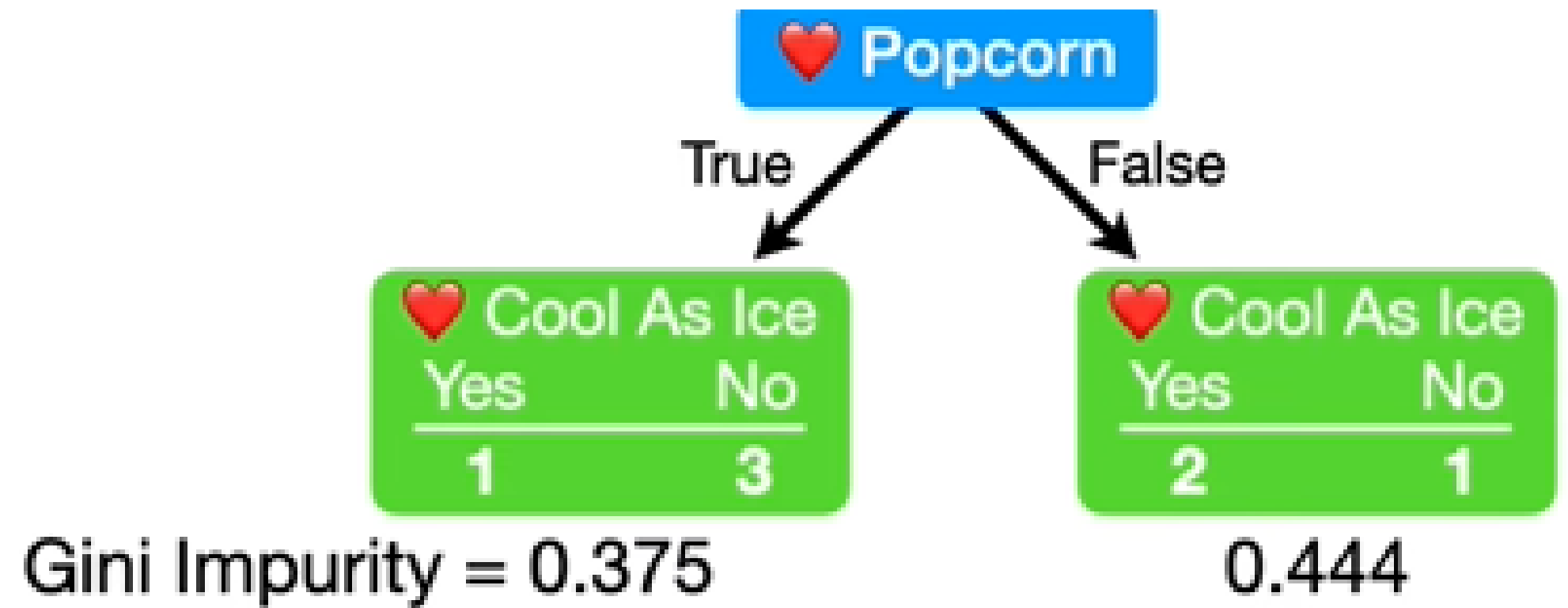
And when we do the math, we get **0.375**.



Departamento de
Computação - **UFSCar**



Impureza de Gini



Total **Gini Impurity** = weighted average of **Gini Impurities** for the **Leaves**

$$= \left(\frac{4}{4+3} \right) 0.375 + \left(\frac{3}{4+3} \right) 0.444$$

= 0.405



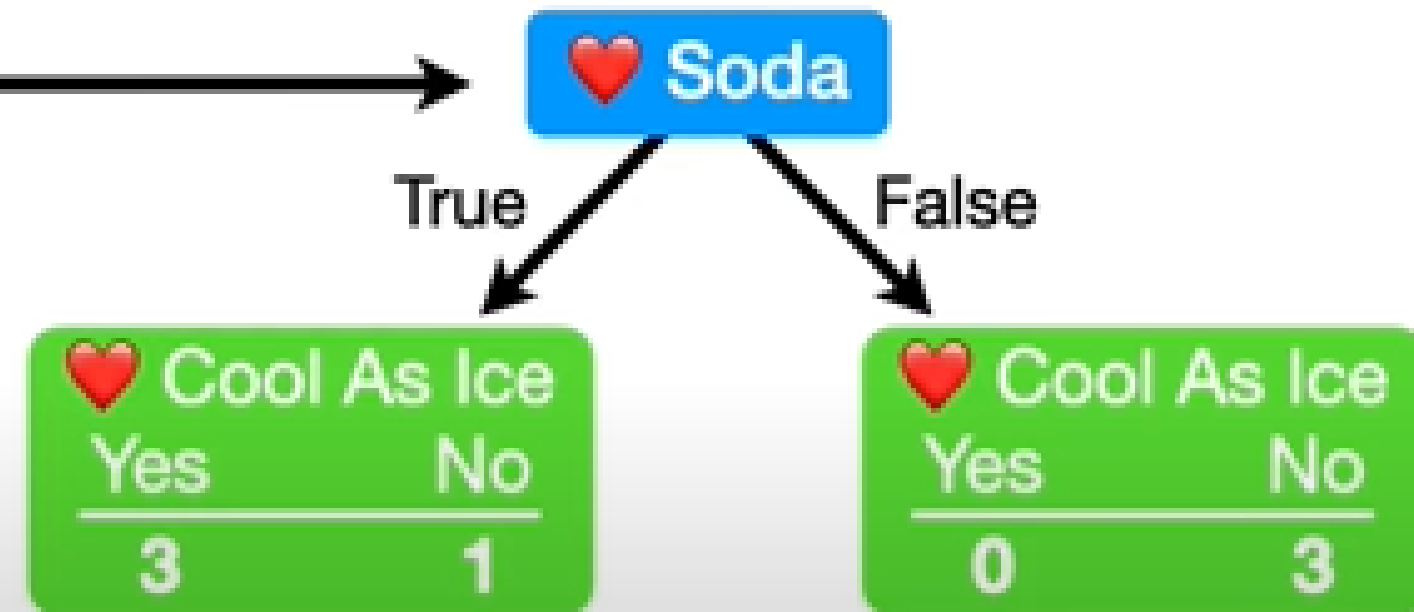
Departamento de
Computação - **UFSCar**



Impureza de Gini

Likewise, the **Gini Impurity** for **Loves Soda** is **0.214**.

Gini Impurity for Loves Soda = 0.214



Departamento de
Computação - **UFSCar**



Impureza de Gini - Valores Numéricos

For example, to calculate the **Gini Impurity** for the first value...

	Age	Loves Cool As Ice
9.5	7	No
15	12	No
26.5	18	Yes
36.5	35	Yes
44	38	Yes
66.5	50	No
	83	No

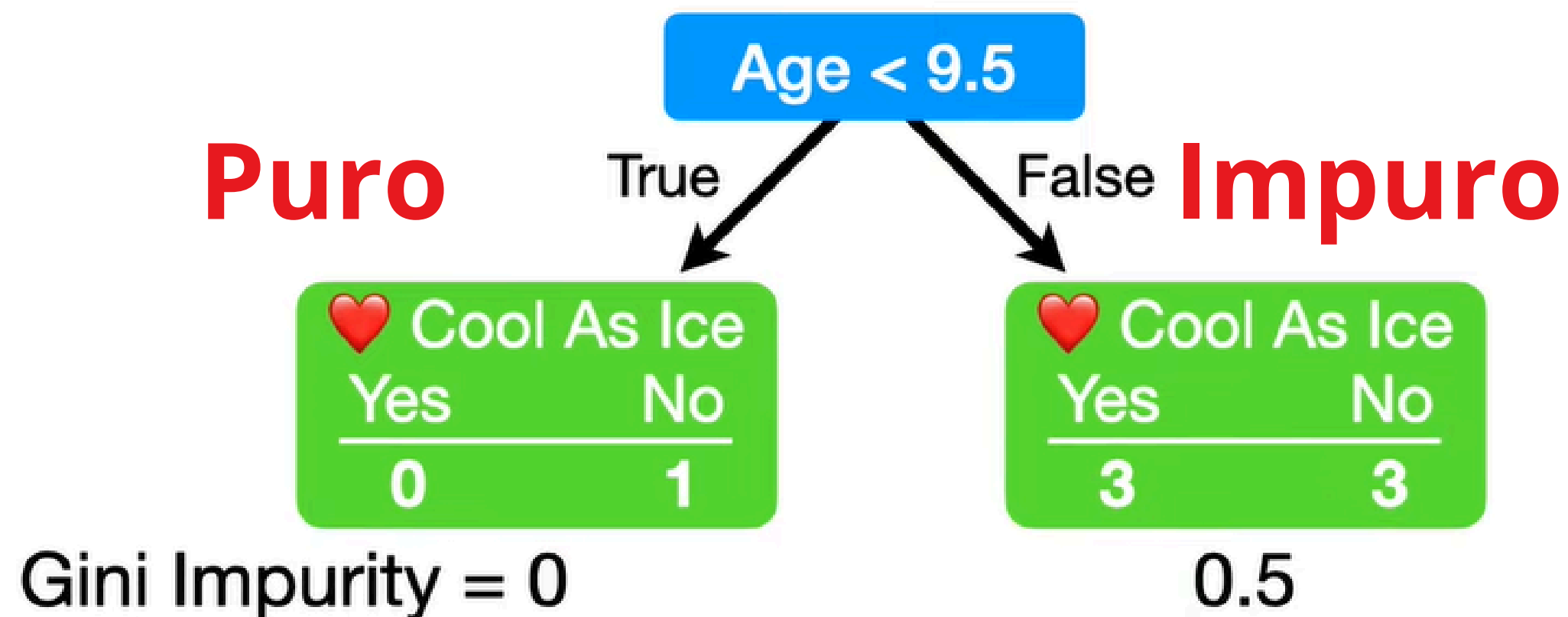
Tiramos a média de idade entre as linhas vizinhas e calculamos a impureza de Gini uma por uma!



Departamento de
Computação - **UFSCar**



Impureza de Gini - Valores Numéricos



$$\text{Total Gini Impurity} = \left(\frac{1}{1+6}\right) 0 + \left(\frac{6}{1+6}\right) 0.5 = 0.429$$



Impureza de Gini - Valores Numéricos



Impureza de Gini

- Impureza de Gini para “Loves PopCorn” = 0.405
- Impureza de Gini para “Loves Soda” = 0.214
- Impureza de Gini para “Age < 15” = 0.214



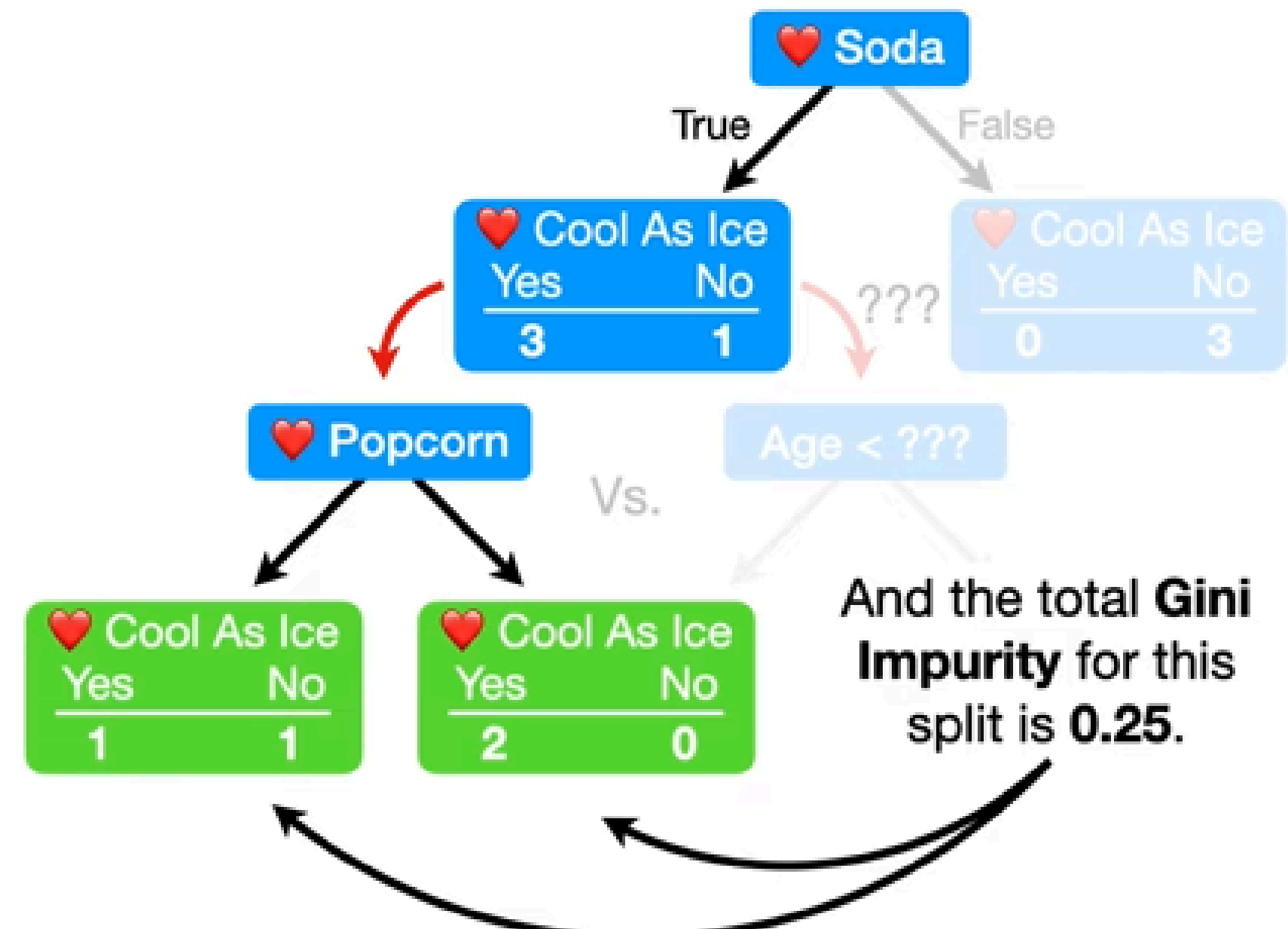
Departamento de
Computação - **UFSCar**



Impureza de Gini



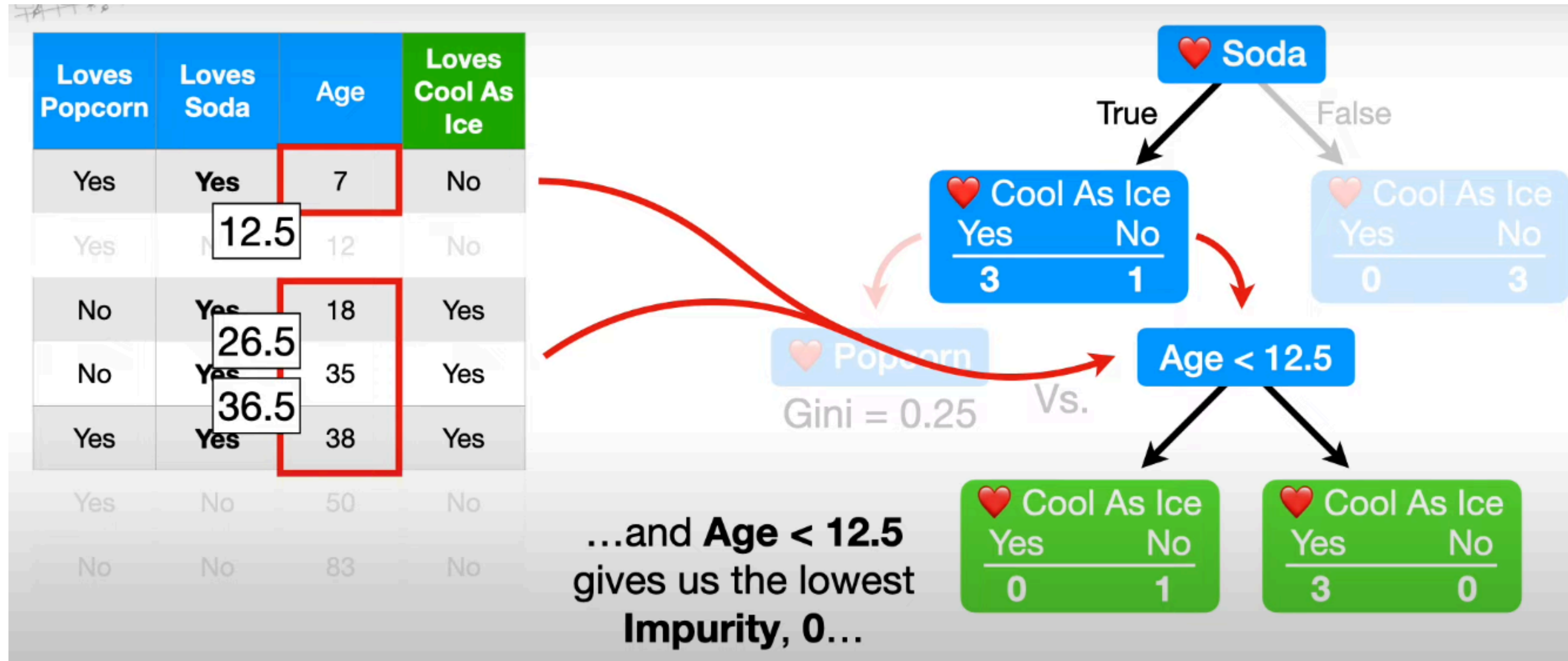
Loves Popcorn	Loves Soda	Age	Loves Cool As Ice
Yes	Yes	7	No
Yes	No	12	No
No	Yes	18	Yes
No	Yes	35	Yes
Yes	Yes	38	Yes
Yes	No	50	No
No	No	83	No



Departamento de
Computação - **UFSCar**



Impureza de Gini



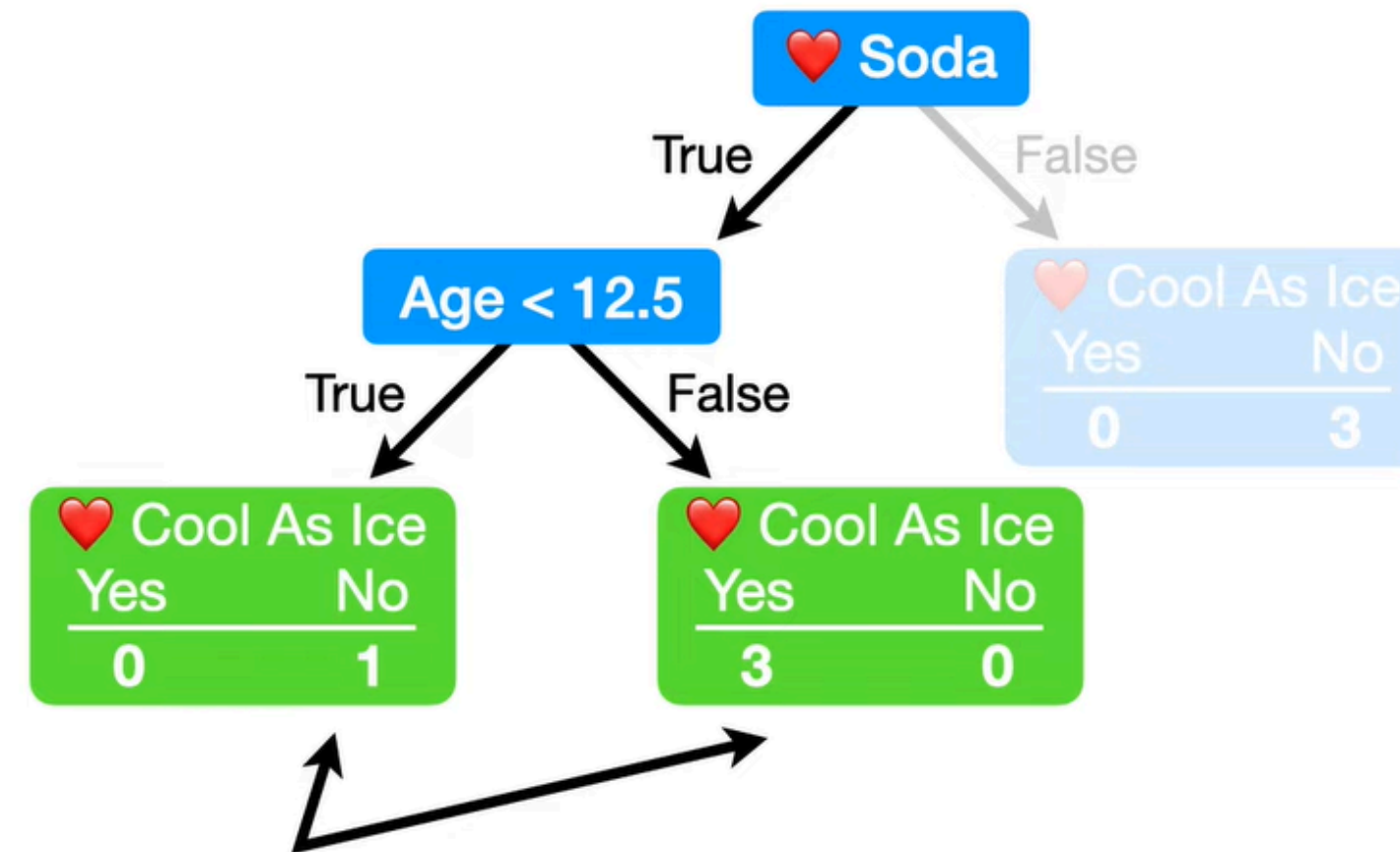
Departamento de
Computação - **UFSCar**



Impureza de Gini



Loves Popcorn	Loves Soda	Age	Loves Cool As Ice
Yes	Yes	7	No
Yes	No	12	No
No	Yes	18	Yes
No	Yes	35	Yes
Yes	Yes	38	Yes
Yes	No	50	No
No	No	83	No



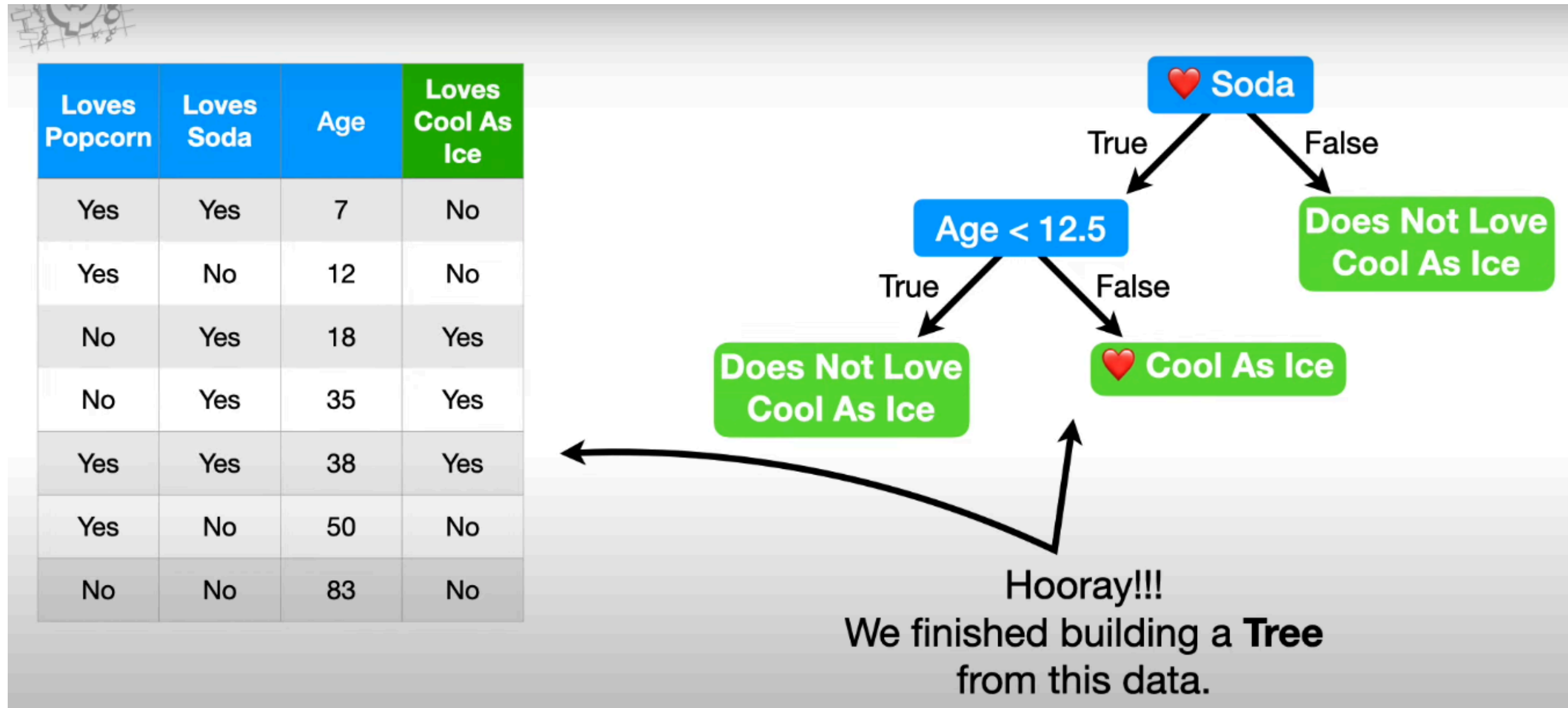
NOTE: These are **Leaves** because there is no reason to continue splitting these people into smaller groups.



Departamento de
Computação - **UFSCar**



Impureza de Gini



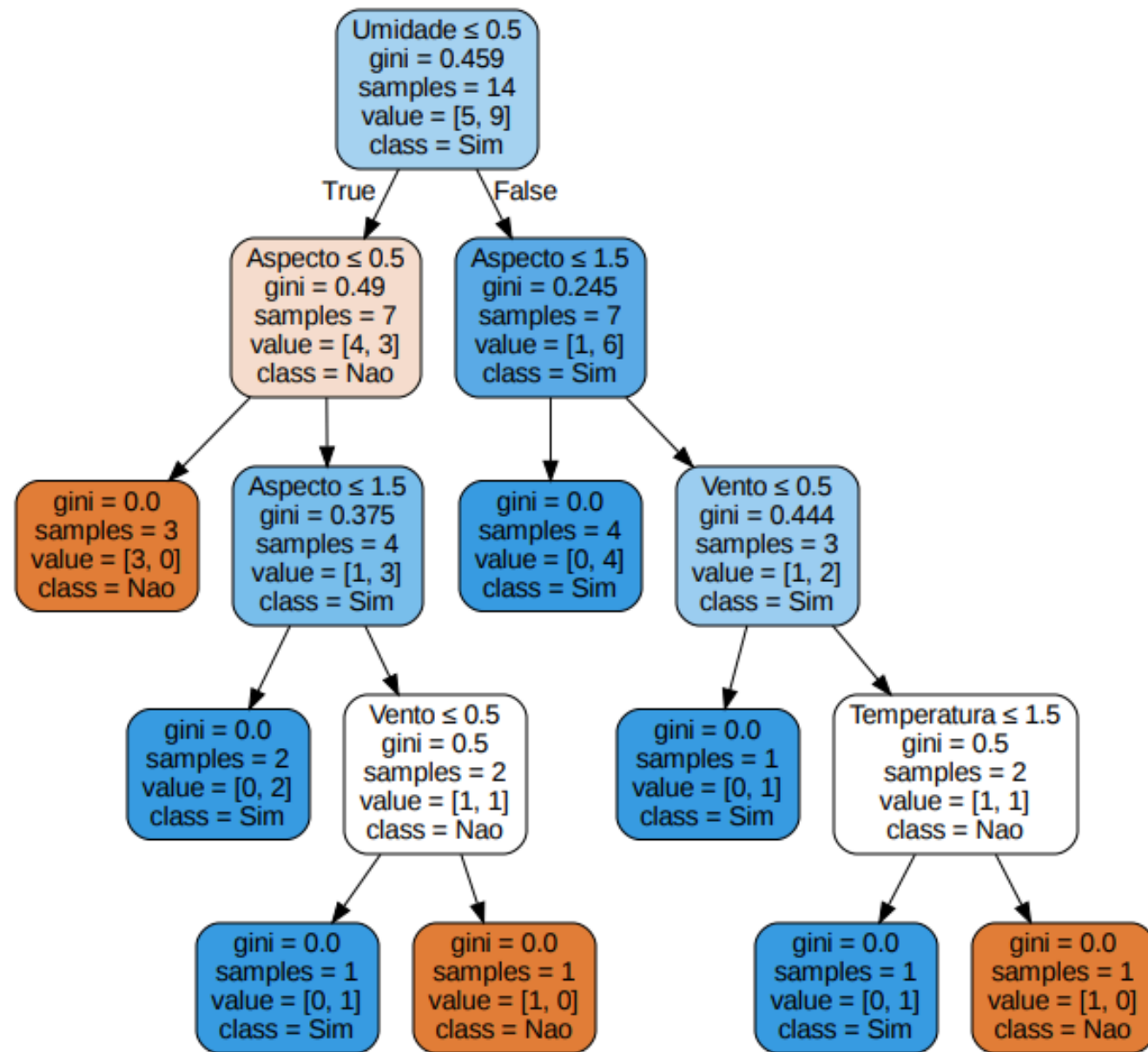
Departamento de
Computação - **UFSCar**



Exemplos de Treino

Dia	Aspecto	Temp.	Humidade	Vento	Jogar Tênis
D1	Sol	Quente	Elevada	Fraco	Não
D2	Sol	Quente	Elevada	Forte	Não
D3	Nuvens	Quente	Elevada	Fraco	Sim
D4	Chuva	Ameno	Elevada	Fraco	Sim
D5	Chuva	Fresco	Normal	Fraco	Sim
D6	Chuva	Fresco	Normal	Forte	Não
D7	Nuvens	Fresco	Normal	Fraco	Sim
D8	Sol	Ameno	Elevada	Fraco	Não
D9	Sol	Fresco	Normal	Fraco	Sim
D10	Chuva	Ameno	Normal	Forte	Sim
D11	Sol	Ameno	Normal	Forte	Sim
D12	Nuvens	Ameno	Elevada	Forte	Sim
D13	Nuvens	Quente	Normal	Fraco	Sim
D14	Chuva	Ameno	Elevada	Forte	Não

Jupyter Notebook - Aplicação de Gini

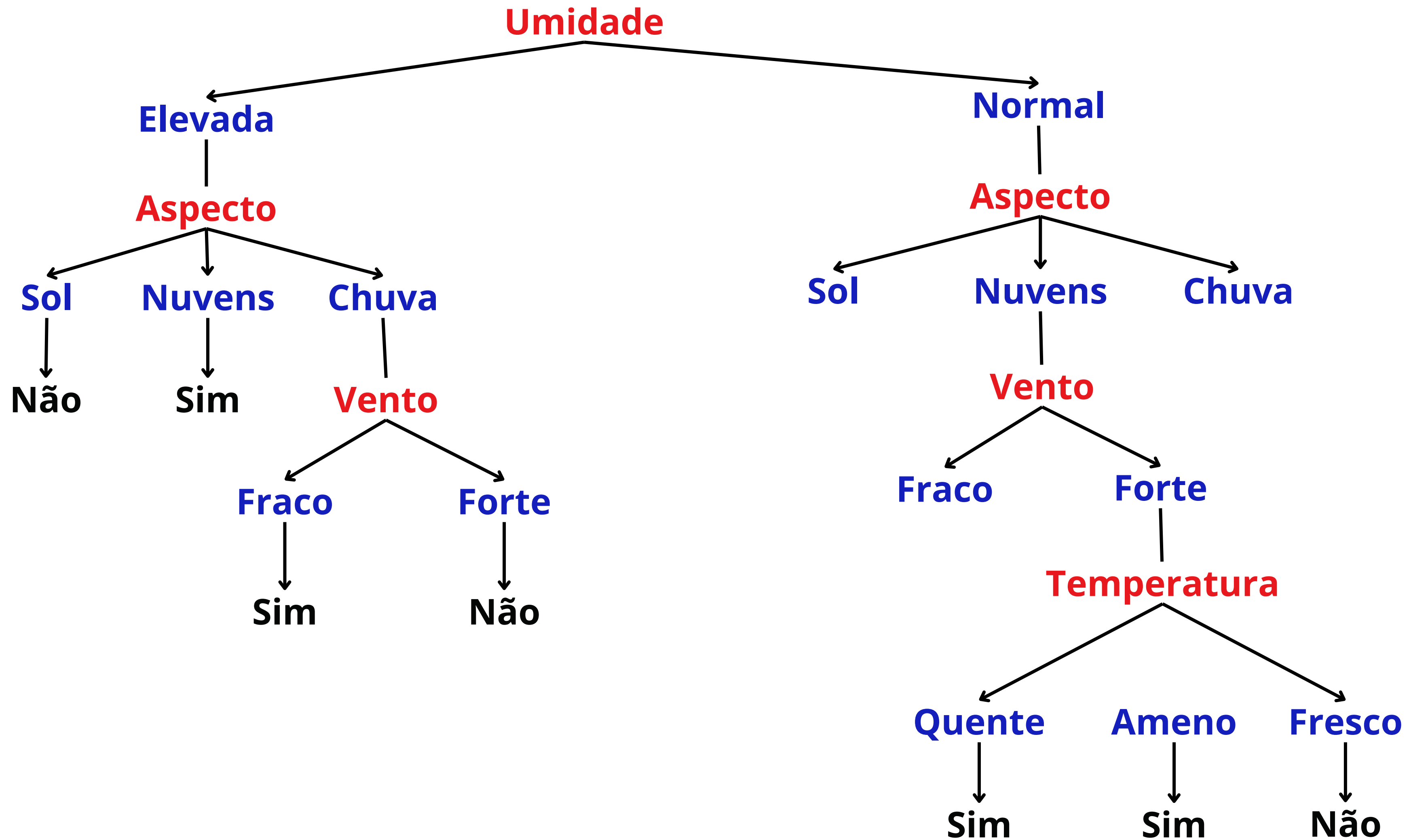


[Clique aqui para acessar o Notebook](#)



Departamento de
Computação - **UFSCar**





Matriz de Confusão - TP, TN, FP, FN

- **Verdadeiro Positivo (TP):** Modelo prevê corretamente a classe positiva.
- **Verdadeiro Negativo (TN):** Modelo prevê corretamente a classe negativa.
- **Falso Positivo (FP):** Modelo prevê incorretamente a classe positiva.
- **Falso Negativo (FN):** Modelo prevê incorretamente a classe negativa.



Departamento de
Computação - **UFSCar**



Matriz de Confusão

Valores Reais

**Valores
Previstos**

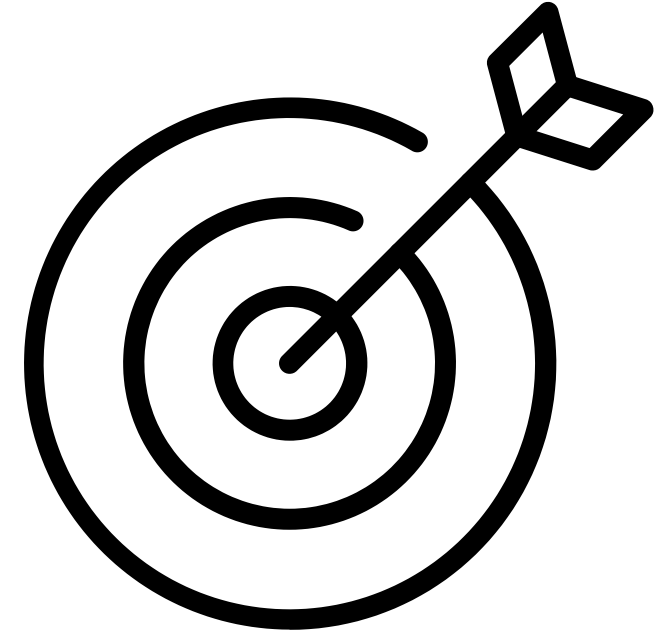
	Positivo 1	Negativo 0
Positivo 1	TP	FP
Negativo 0	FN	TN



Departamento de
Computação - **UFSCar**



Acurácia



$$\text{Acurácia} = \frac{\text{Qtd Predições Corretas}}{\text{Qtd Total de Predições}}$$

$$\text{Acurácia} = \frac{TP + TN}{TP + TN + FP + FN}$$



Departamento de
Computação - **UFSCar**



Classificação - Árvore de Decisão Aplicada em Diagnóstico de Doenças Cardiovasculares

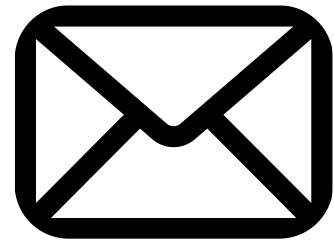
- [Link do Google Colab](#)
- [Teoria de Decision Tree bem explicada - StatQuest](#) (*USADO NESTES SLIDES*)



Departamento de
Computação - **UFSCar**



Contato



leonardorossi@estudante.ufscar.br



Léo Rossi Dourado



leorossi07



Departamento de
Computação - **UFSCar**

