

Universidade Federal do Pará  
Programa de Pós-graduação em Engenharia de Processos  
(PPGEP)  
Disciplina: Tópicos Especiais em Engenharia de Processos  
Prof. Dr. Alan de Souza  
Aula 1

Setembro/2025

## Apresentação do professor

- Graduação em engenharia da computação
- Especialização em gerência e projetos de software
- Mestrado e doutorado na área de inteligência artificial na UFPA, projeto com dados da Albras (PTG e processo de produção do alumínio primário)
- Experiência como analista de sistemas
- Professor universitário desde de 2014
- Atualmente trabalha na UEPA/DMEI
- Em termos de pesquisa, tem atuado bastante nas áreas:
  - Machine Learning, Processamento de Linguagem Natural (NLP), Ciência de Dados, Redes Sociais (Twitter, Youtube), avaliações do Tripadvisor, Google Play.

## Apresentação da disciplina

- Nome: Tópicos especiais em engenharia de processos - foco em técnicas de machine learning / aprendizado de máquina
- Período e Horário:
  - Aulas: de 29/set a 03/10 de 2025 (exceto quarta-feira, 01/10)
  - Avaliações: a partir de 03/10 de 2025
- Metodologia: mão na massa (*hands on*), D.I.Y. (*do it yourself*), ABP (aprendizado baseado em projetos) - *Project-Based Learning* (PBL)
  - Uso de notebook durante as aulas,
  - Linguagem de programação Python

## Apresentação da disciplina

- Avaliação: os discentes deverão desenvolver um projeto aplicando as técnicas de *machine learning* explicadas durante as aulas e deverão entregar:
  - um relatório e
  - o código-fonte desenvolvido.

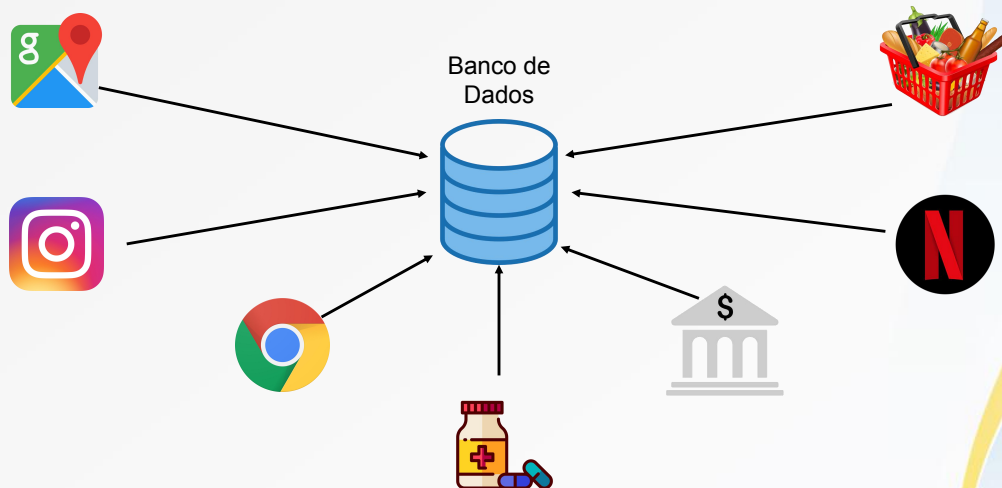
## Reconhecimento da turma

- Qual a sua graduação?
- Lembra de conceitos de Estatística?
- Tem domínio de programação de computadores?
- Qual é o tema da sua dissertação de mestrado?
- Já trabalha na área de formação?

## Introdução

- Quando se fala em dados, qual a primeira coisa que vem na sua cabeça?
- Por que dados são importantes? Por que coletar dados?
- Que tipo de uso de dados você conhece?

## Introdução - coleta de dados



## Introdução - possibilidades em dados

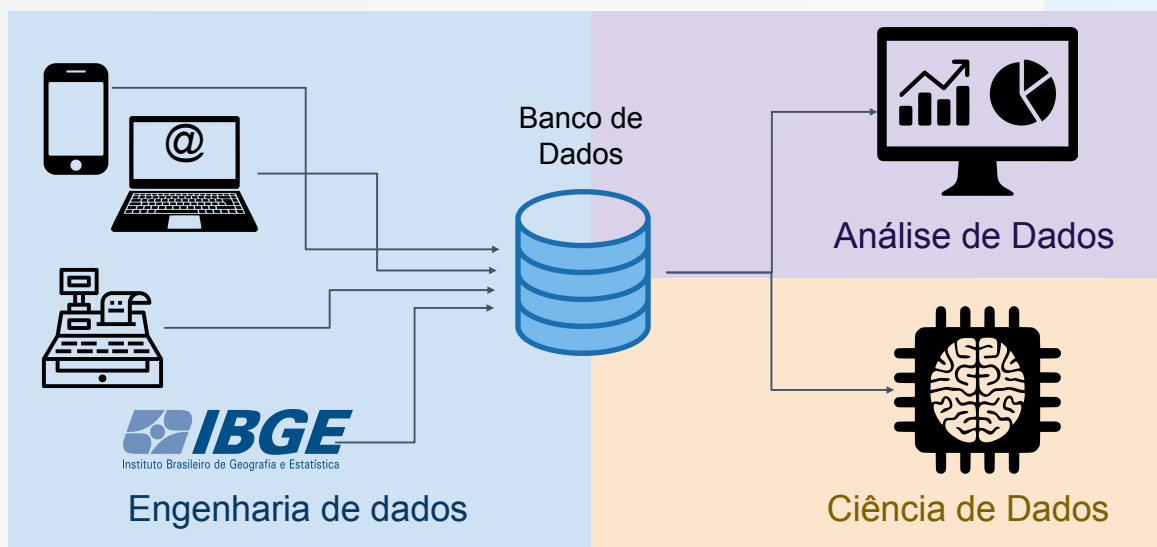
- Previsão de vendas
- Tempo de garantia
- Limite de crédito
- Manutenção de equipamento industrial
- Recomendação de produtos
- Definição de metas
- Processamento de Linguagem Natural
- Conhecer diferentes grupos de clientes
- Validar produtos com usuários
- Entender sazonalidade
- Visão computacional
- Performance de atletas
- Eficácia de vacinas e medicamentos

**Tomada de decisões mais certas**

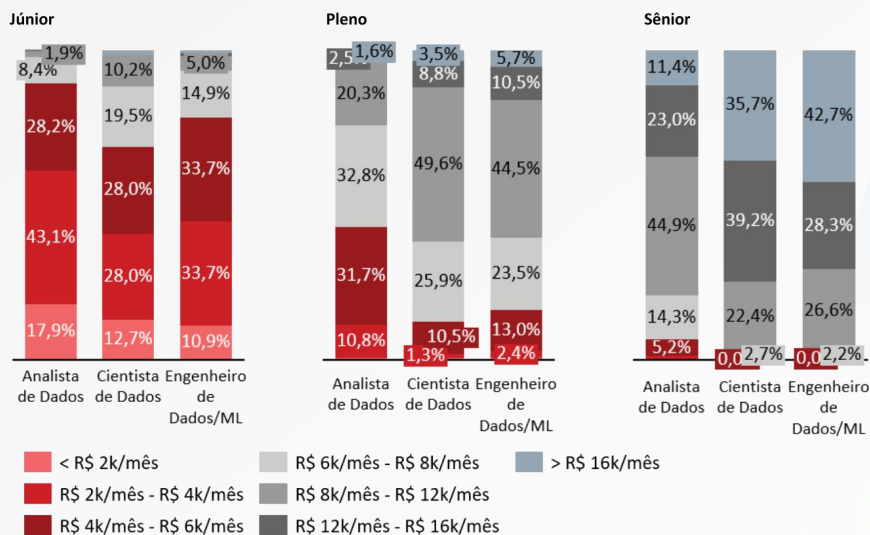
## Introdução - profissionais de dados (básico)

- **Engenheiro de dados:** responsável por coletar e tratar dados de diferentes fontes e disponibilizá-los para que outras pessoas possam consumir esses dados (facilitar o acesso: formato de planilha, dashboard, Power BI)
- **Analista de dados:** responsável por consumir os dados e dar uma visão mais orientada ao negócio; monitorar os indicadores; tomada de decisões.
- **Cientista de dados:** também consome os dados, mas tenta prever os próximos acontecimentos através de dados históricos.

## Introdução - profissionais de dados



## Introdução - profissionais de dados (salários)



Pesquisa State of Data Brasil 2024

# DÚVIDAS?

## Introdução - o que é *machine learning*?

### Qual é a fruta?

Redonda/esfera

Suculenta

Vermelha

Doce



### Qual é a fruta?

| Redondada | Suculenta | Vermelha | Doce | Fruta   |
|-----------|-----------|----------|------|---------|
| 0         | 1         | 1        | 1    | Morango |
| 1         | 0         | 0        | 0    | Limão   |
| 1         | 1         | 0        | 1    | Pera    |
| 0         | 0         | 0        | 1    | Banana  |
| 1         | 1         | 1        | 1    | Cereja  |
| 1         | 1         | 1        | 0    | Tomate  |
| 1         | 1         | 1        | 1    | Maçã    |

Atributos  
Covariáveis  
Variáveis explicativas/preditores/independentes

Alvo  
Resposta  
Dependente

**Algoritmos de aprendizado de máquina “descobrem” regras com base em exemplos.**

**Só “conhecem” objetos, eventos, etc, que foram apresentados a ele, com uma probabilidade associada.**

## Introdução - o que é *machine learning*?

O aprendizado automático (...) é um subcampo da Engenharia e da Ciência da Computação que evoluiu do estudo de **reconhecimento de padrões** e da teoria do aprendizado computacional em **inteligência artificial** [1].

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" [2]

[1] <https://www.britannica.com/technology/machine-learning>

[2] [https://books.google.com.br/books?id=Dn-Gdoh66sgC&pg=PA89&redir\\_esc=y#v=onepage&q&f=false](https://books.google.com.br/books?id=Dn-Gdoh66sgC&pg=PA89&redir_esc=y#v=onepage&q&f=false)

## Como separar essas cervejas pelos seus atributos?

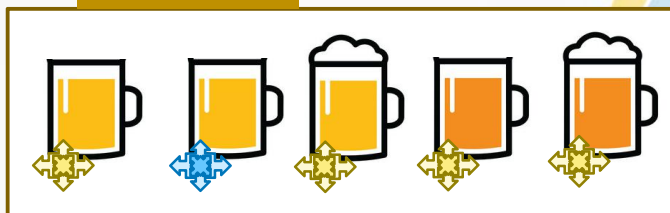
Pale-Ale



Weissbier

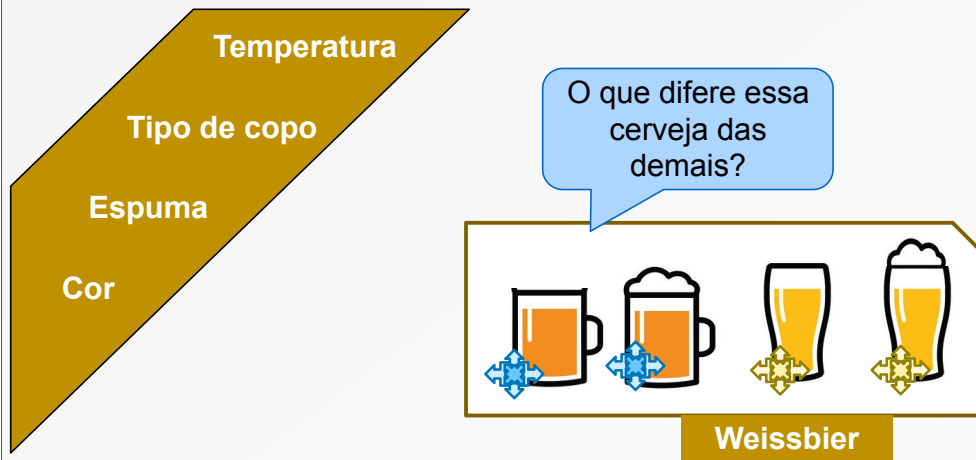


Pilsen

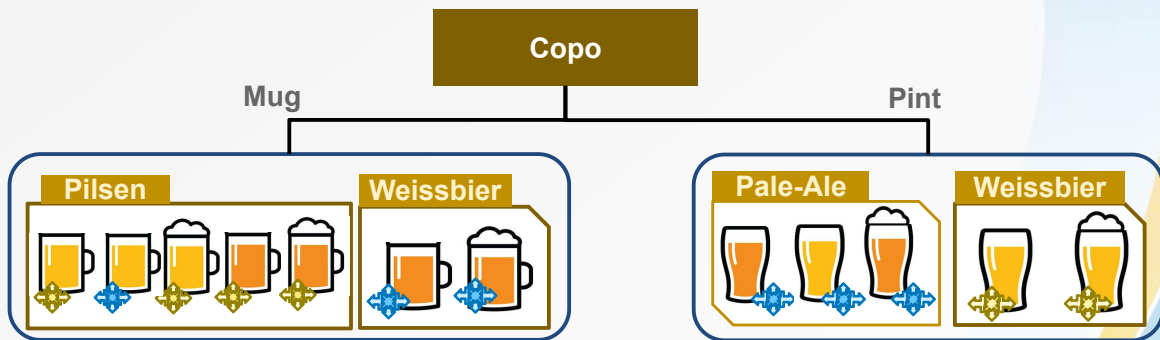




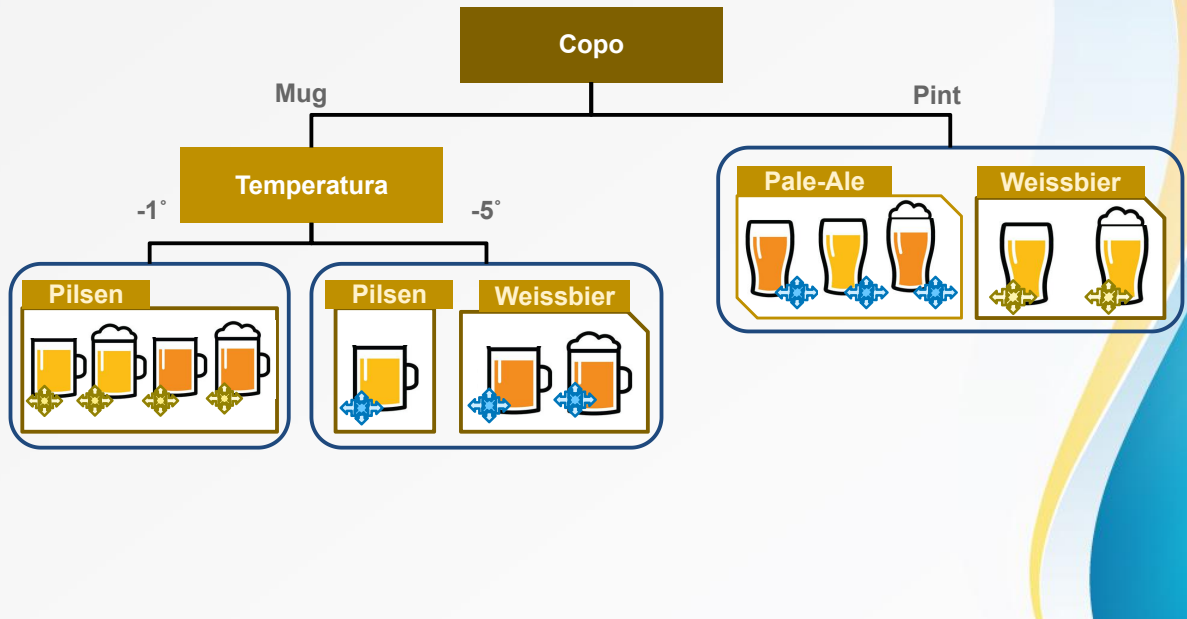
## Quais atributos temos?



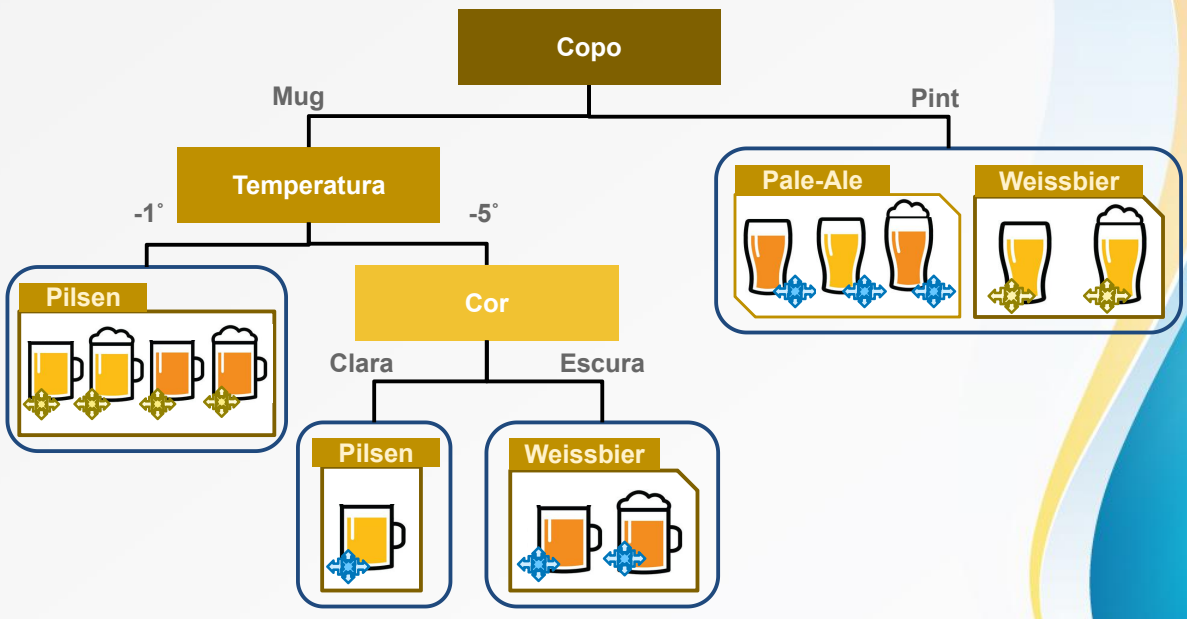
## Criando regras



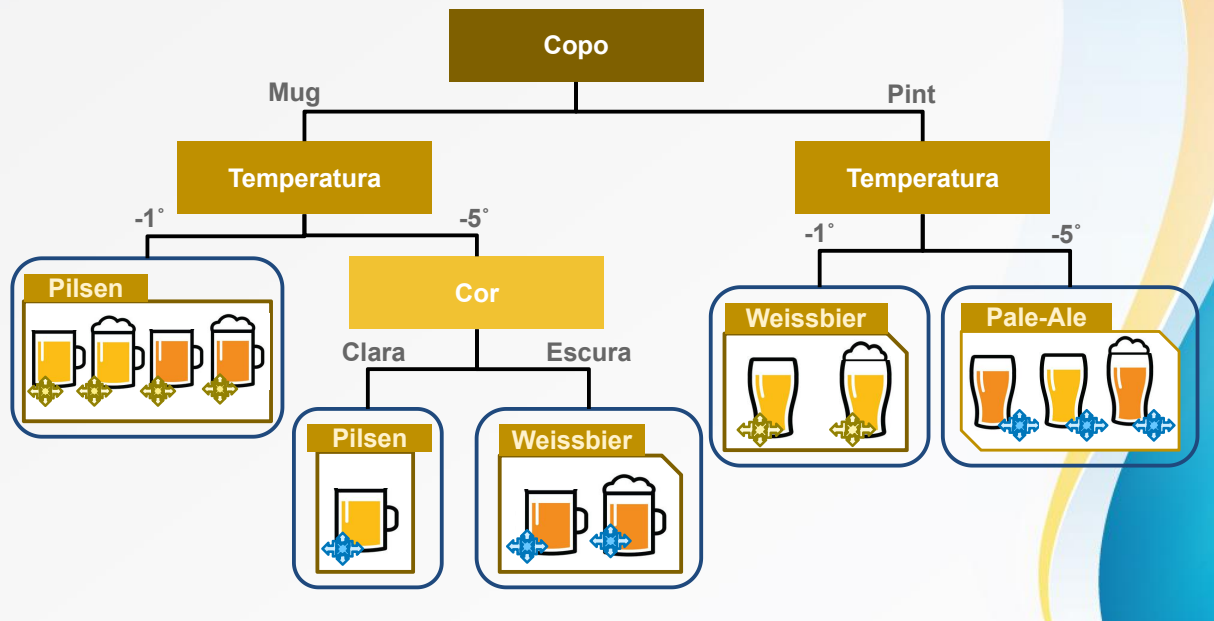
## Criando regras



## Criando regras



## Criando regras



## O que acabamos de fazer?

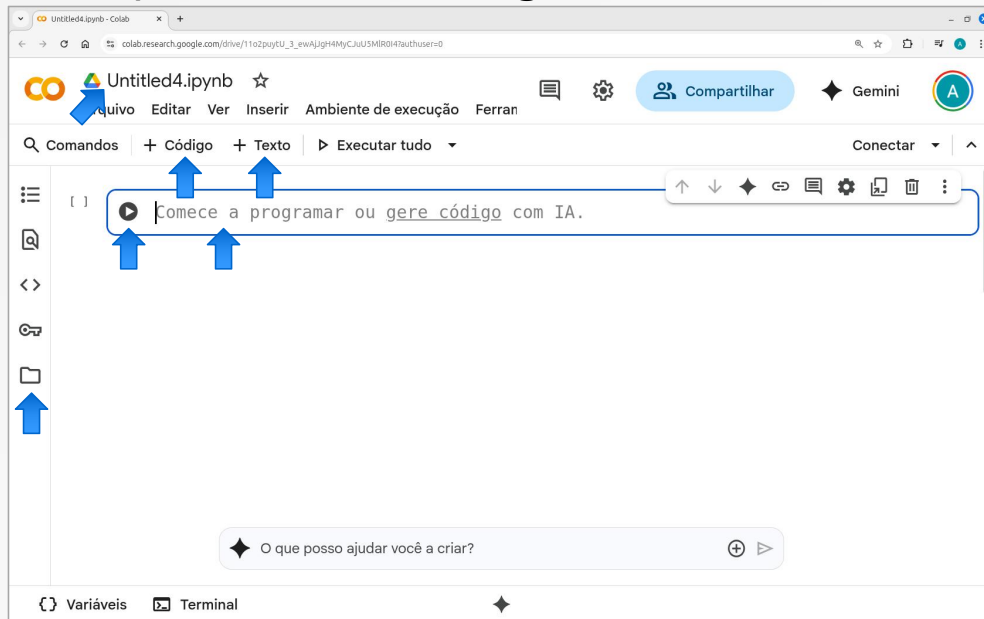
- Acabamos de usar o algoritmo Árvore de Decisão para criar um **modelo** capaz de separar/classificar as três cervejas!
  - Algoritmo exaustivo, porque chega ao resultado por tentativa e erro.
  - A espuma foi relevante na classificação?
  - A árvore fatia o conjunto de dados.
  - Algoritmo muito usado em outros algoritmos.
- Não precisamos programar regra por regra.

## Tabela de dados

| id | copo | temperatura | cor    | espuma | classe    |
|----|------|-------------|--------|--------|-----------|
| 1  | mug  | -5          | escura | não    | weissbier |
| 2  | mug  | -5          | escura | sim    | weissbier |
| 3  | pint | -1          | clara  | não    | weissbier |
| 4  | pint | -1          | clara  | sim    | weissbier |
| 5  | pint | -5          | escura | não    | pale-ale  |
| 6  | pint | -5          | clara  | não    | pale-ale  |
| 7  | pint | -5          | escura | sim    | pale-ale  |
| 8  | mug  | -1          | clara  | não    | pilsen    |
| 9  | mug  | -5          | clara  | não    | pilsen    |
| 10 | mug  | -1          | clara  | sim    | pilsen    |
| 11 | mug  | -1          | escura | não    | pilsen    |
| 12 | mug  | -1          | escura | sim    | pilsen    |

# DÚVIDAS?

## Parte prática com o Google Colab



## Projeto prático - Frutas

### # Importação dos dados:

```
import pandas as pd
```

```
df = pd.read_excel("dados_frutas.xlsx")
df
```

## Projeto prático - Frutas

# Separação de dados de entrada (X) e de saída (y) do modelo:

```
y = df['Fruta']
```

```
entradas = ['Arredondada', 'Suculenta', 'Vermelha', 'Doce']
```

```
X = df[entradas]
```

## Projeto prático - Frutas

# Criação e treino do modelo:

```
from sklearn import tree
```

```
arvore = tree.DecisionTreeClassifier(random_state=42)
```

```
arvore.fit(X, y)
```

# Uso do modelo para fazer predição:

```
arvore.predict([[0, 0, 0, 0]])
```

## Projeto prático - Frutas

### # Exibição do desenho da árvore de decisão:

```
import matplotlib.pyplot as plt
from sklearn import tree
plt.figure(dpi=400, figsize=[4,4])
tree.plot_tree(arvore,
               feature_names=entradas,
               class_names=arvore.classes_,
               filled=True)
plt.show()
```

## Projeto prático - Frutas

### # Exibição das probabilidades de predição de cada classe:

```
proba = arvore.predict_proba([[1,1,1,1]])[0]
pd.Series(proba, index=arvore.classes_)
```

# DÚVIDAS?

## Exercício

- 1) Crie a árvore de decisão para os dados de *Cerveja*.
  - a) Importação dos dados;
  - b) Separação dos dados em entradas e saída;
  - c) Criação e treino do modelo;
  - d) Pelo menos uma predição qualquer;
  - e) Exibição do desenho da árvore;
  - f) Exibição das probabilidades de pelo menos uma predição qualquer.



## Projeto prático - Cerveja

### # Importação dos dados:

```
import pandas as pd
```

```
df = pd.read_excel("dados_cerveja.xlsx")  
df.head()
```

## Projeto prático - Cerveja

### # Separação de dados de entrada (X) e de saída (y) do modelo:

```
entradas = ['temperatura','copo','espuma','cor']
```

```
saida = 'classe'
```

```
X = df[entradas]
```

```
y = df[saida]
```

```
X = X.replace({
```

```
    "mug": 1, "pint":2,
```

```
    "sim":1, "não":0,
```

```
    "clara": 0, "escura":1,
```

```
})
```

## Projeto prático - Cerveja

### # Criação e treino do modelo:

```
from sklearn import tree
arvore = tree.DecisionTreeClassifier()
arvore.fit(X, y)
```

### # Uso do modelo para fazer predição:

```
arvore.predict([[-5, 2, 1, 0]])
```

## Projeto prático - Cerveja

### # Exibição do desenho da árvore de decisão:

```
import matplotlib.pyplot as plt
from sklearn import tree
plt.figure(dpi=400)
tree.plot_tree(arvore,
               feature_names=entradas,
               class_names=arvore.classes_,
               filled=True)
plt.show()
```

## Projeto prático - Cerveja

# Exibição das probabilidades de predição de cada classe:

```
proba = arvore.predict_proba([[-1, 1, 0, 0]])[0]  
pd.Series(proba, index=arvore.classes_)
```

# DÚVIDAS?