

Sobre o Dataset Iris

O **Iris Dataset** é um dos conjuntos de dados mais clássicos da ciência de dados, introduzido por **Ronald Fisher** em 1936.

- **Total de observações:** 150 flores
- **Classes (espécies):**
 - **setosa**
 - **versicolor**
 - **virginica**
- **Variáveis (features):**
 - Comprimento da sépala (**sepal length**)
 - Largura da sépala (**sepal width**)
 - Comprimento da pétala (**petal length**)
 - Largura da pétala (**petal width**)

Cada linha representa uma flor, com suas medidas e a espécie à qual pertence.

Tipo de Problema: Classificação Multinomial

Este é um problema de **classificação**, onde o objetivo é prever a espécie da flor com base nas 4 medidas.

- ☒ **Tipo de problema:** Classificação
- ☒ **Número de classes:** 3
- ☒ **Tipo de classificação:** **Multinomial** (ou multiclasse)
- ☒ **Algoritmos possíveis:** regressão logística, k-vizinhos, árvores de decisão, etc.

Se estivéssemos lidando com apenas **duas espécies**, o problema seria **binário**.

Exemplo de Aplicação

- **X (variáveis explicativas):** medidas das pétalas e sépalas
- **y (variável alvo):** espécie da flor (**setosa**, **versicolor**, **virginica**)
- **Objetivo:** Construir um modelo que classifique corretamente a espécie com base nas medidas