Iris Dataset Background

- The Iris dataset is one of the most famous datasets in statistics and machine learning.
- It was introduced in 1936 by Ronald A. Fisher, a British statistician.
- The dataset was originally used in his paper:
"The use of multiple measurements in taxonomic problems".
- Fisher collected data about three species of Iris flowers:
* Iris setosa
* Iris versicolor
* Iris virginica
- Each flower sample has 4 measurements (features):
1. Sepal length
2. Sepal width
3. Petal length
4. Petal width
- There are 150 samples total (50 per species).
- The dataset was originally collected from the Gaspé Peninsula in Canada.
- Why it matters:
* It's small and clean — easy to practice on.
* It's historically important — used in early examples of discriminant analysis.
* It's widely available in scikit-learn and other libraries.