

Practical 2 - ML (Iris Dataset)

Monday, 26 January 2026

4:37 PM

The **Iris dataset** is a **small, famous dataset** used to teach **Machine Learning basics**, especially **classification**.

It contains measurements of **iris flowers** and the task is to **identify the flower species** based on those measurements.

What Data Does It Contain?

The dataset has **150 flower samples**.

Flower Species (Target / Output)

There are **3 classes**:

1. **Setosa**
2. **Versicolor**
3. **Virginica**

Features (Input Data)

Each flower has **4 measured features**:

Feature	Meaning
Sepal Length (cm)	Length of the outer leaf
Sepal Width (cm)	Width of the outer leaf
Petal Length (cm)	Length of inner colorful leaf
Petal Width (cm)	Width of inner colorful leaf

Dataset Structure (ML View)

- **Rows (150)** → Each row = one flower
- **Columns (4 + 1)**
 - 4 input features
 - 1 output label (species)

$X = [\text{sepal_length}, \text{sepal_width}, \text{petal_length}, \text{petal_width}]$

$y = \text{species}$