Iris dataset: source and safe-use guide

Notes for SHS | Data Science | Topic 87 | 25 November 2026

Original authoritative source

Scikit-Learn built-in Iris dataset: https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_iris.html

150 Iris flower measurements, 4 numerical features in centimetres; setosa, versicolor and virginica (50 each). From the historical Fisher Iris benchmark; inspect iris.DESCR for source attribution.

Access using Python `from sklearn.datasets import load_iris; iris=load_iris(as_frame=True)`. No external download or private student data is required. This package includes reproducible code rather than redistributing a separate third-party raw dataset file.

Limitations and ethics

A compact historical benchmark with balanced labels is not proof of general real-world accuracy, probability calibration or fairness. A four-measurement classifier does not classify arbitrary flower photographs.

Reference

Scikit-Learn documentation: https://scikit-learn.org/stable/datasets/toy_dataset.html