

# Supervised Machine Learning – Iris Flower Classification

## 1. Introduction

A botanical research centre aims to automate the identification of Iris flower species based on physical measurements. Manual identification is time-consuming, error-prone, and not scalable. This project implements supervised machine learning algorithms to classify Iris flowers automatically.

## 2. Dataset Description

The Iris dataset contains 150 samples equally distributed among three classes: Setosa, Versicolor, and Virginica. Each sample has four numerical features: Sepal Length, Sepal Width, Petal Length, and Petal Width.

## 3. Data Preprocessing

- Removed unnecessary columns
- Encoded categorical target labels
- Split data into 50% training and 100% testing
- Standardized features where required

## 4. Models Implemented

- K-Nearest Neighbours (KNN)
- Logistic Regression
- Naive Bayes

## 5. Evaluation Metrics

Models were evaluated using accuracy, precision, recall, F1-score, and confusion matrices.

## 6. Results

KNN achieved the highest accuracy, followed by Logistic Regression and Naive Bayes. All models performed well due to the clean and balanced nature of the dataset.

## 7. Conclusion

This project demonstrates the effectiveness of supervised learning algorithms for classification problems. KNN was selected as the best model based on performance metrics.

## 8. References

- Scikit-learn Documentation
- Fisher, R. A. (1936). *The use of multiple measurements in taxonomic problems*.