

Iris Flower Species Classification using Machine Learning

Objective

The objective of this project is to build, train, and evaluate machine learning models that can accurately classify Iris flower species based on their physical measurements. This project demonstrates the application of supervised learning techniques in biological data classification.

Dataset Description

The Iris dataset is a well-known benchmark dataset available in the Scikit-learn library. It contains 150 records of Iris flowers divided equally among three species: Iris-setosa, Iris-versicolor, and Iris-virginica. Each record consists of four numerical features: sepal length, sepal width, petal length, and petal width (in centimeters).

Methodology

1. The dataset was loaded using the Scikit-learn library.
2. Exploratory data analysis was performed to understand feature distributions.
3. The dataset was split into 80% training data and 20% testing data.
4. Three machine learning classification models were trained: Logistic Regression, K-Nearest Neighbors, and Decision Tree Classifier.
5. Models were evaluated using accuracy, confusion matrix, and classification report metrics.

Results

Among the trained models, Logistic Regression achieved the highest accuracy of approximately 97%. K-Nearest Neighbors also performed well with slightly lower accuracy, while the Decision Tree model showed comparatively lower performance due to overfitting tendencies.

Conclusion

This project demonstrates that machine learning models can effectively classify biological data with high accuracy when relevant features are used. The Iris dataset serves as an excellent example of how supervised learning techniques can be applied to real-world classification problems. Such models can be extended to other domains such as agriculture, medicine, and environmental studies.