

TECHNICAL REPORT: IRIS FLOWER SPECIES PREDICTION

1. INTRODUCTION

The Iris flower dataset is one of the most renowned datasets in machine learning, widely used for classification exercises. It contains morphological measurements of Iris flowers, sepal length, sepal width, petal length, and petal width, across three species: *Setosa*, *Versicolor*, and *Virginica*. The goal of this project is to build and deploy a machine learning model capable of accurately classifying Iris species from these measurements.

2. DATASET DESCRIPTION

- **Samples:** 150 Iris flowers
- **Features:**
 - Sepal length (cm)
 - Sepal width (cm)
 - Petal length (cm)
 - Petal width (cm)
- **Target:** Iris species (*Setosa*, *Versicolor*, *Virginica*)

The dataset is loaded using `sklearn.datasets.load_iris` and managed as a Pandas DataFrame.

3. METHODOLOGY

Step 1: Data Loading

The Iris dataset is loaded from the sklearn library and organized for analysis.

Step 2: Data Splitting

The dataset is split into 80% training and 20% testing subsets, with a fixed random seed for reproducibility.

Step 3: Model Training

A Logistic Regression classifier is trained on the training data, chosen for its simplicity and proven effectiveness in multi-class classification.

Step 4: Prediction

The model predicts the species on the unseen test data.

Step 5: Evaluation

Model performance is assessed using:

- **Accuracy Score:** Proportion of correct predictions on the test set.
- **Confusion Matrix:** Detailed class-wise prediction accuracy, identifying misclassifications.

Step 6: Interactive User Interface

To enhance usability, the model is deployed as a web app using Streamlit. Users can input custom flower measurements interactively through sliders and get instant species predictions, along with model accuracy and confusion matrix visualizations.

4. RESULTS

- **Accuracy:** The Logistic Regression model achieved an accuracy of approximately **97%** on the test data.
- **Confusion Matrix:** Minimal misclassifications were observed, primarily between *Versicolor* and *Virginica*, which have overlapping feature ranges.
- **USER Interaction:** The Streamlit app provides a user-friendly interface for inputting flower measurements and instantly obtaining species predictions, accompanied by performance insights.

5. CONCLUSION

This project demonstrates the effectiveness of Logistic Regression for Iris species classification based on morphological features. The interactive web app further extends the model's accessibility, enabling users to experiment with their own inputs and understand the model's performance.

6. FUTURE WORK

- Experiment with advanced models such as Random Forest, Support Vector Machines, or Neural Networks.
- Implement cross-validation for more robust performance assessment.
- Expand the user interface with additional features, including graphical representations or mobile compatibility.
- Incorporate larger and more diverse datasets for enhanced generalization.