Project Title: Iris flower Classification

Student Name: Shahid Siddiqui

Roll No.: 03

Branch: CSE AI

Aktu roll no. : 202401100300223

Date: 11/03/2025

Introduction

The Iris flower classification is a well-known problem in machine learning and pattern recognition. It involves categorizing iris flowers into one of three species: Setosa, Versicolor, and Virginica, based on four features:

Sepal Length

Sepal Width

Petal Length

Petal Width

Dataset

The dataset used for classification is the Iris dataset, which contains 150 samples, evenly distributed across the three species. Each sample is characterized by four numerical features.

Methodology

1.

2.

3.

4.

Data Preprocessing

Load the dataset

Handle missing values (if any)

Normalize the features for better performance

Exploratory Data Analysis (EDA)

Visualization of feature distribution using histograms and scatter plots

Correlation analysis between features

Model Selection

Dierent classification algorithms are considered:

Logistic Regression

K-Nearest Neighbors (KNN)

Support Vector Machine (SVM)

Decision Tree

Random Forest

Split data into training (80%) and testing (20%) sets

Model Training and Evaluation

Train models using the training dataset

Evaluate models using accuracy, precision, recall, and F1-score

Use confusion matrices to analyze misclassification

Results

The Random Forest classifier achieved the highest accuracy of around 97%.

SVM also performed well with an accuracy of approximately 96%.

Setosa was the easiest species to classify, while some misclassification occurred between

Versicolor and Virginica.

Conclusion
The Iris dataset is a great introductory problem for classification tasks in machine learning.

The Random Forest model provided the best accuracy, but simpler models like Logistic

Regression and SVM also performed well. Further improvements can be made by fine-tuning

hyperparameters and testing with additional feature engineering techniques.

Future Work

Implement deep learning models for classification.

Utilize advanced feature selection techniques.

Experiment with ensemble learning methods for better accuracy.