# Iris Flower Classification using Machine Learning

## 1. Problem Statement:-

Accurate classification of flower species based on physical measurements is an important task in botanical research and machine learning applications. Manual identification of iris species can be time-consuming and error-prone due to similarities in measurements.

**Objective:**

– To build a Machine Learning classification model that can

– Learn from iris flower measurements

– Accurately classify flowers into Setosa, Versicolor, or Virginica

## 2. Dataset Description:-

– Source: Iris Flower Dataset

– Records: 150 flower samples

– Data Type: Structured (CSV)

**Features:**

– Sepal Length

– Sepal Width

– Petal Length

– Petal Width

**Target Variable:**

– Species (Setosa, Versicolor, Virginica)

## 3. Data Preprocessing & Feature Engineering:-

– Loaded dataset from CSV file

– Checked for missing and duplicate values

– Encoded categorical target labels using Label Encoding

– Split dataset into training and testing sets

– Scaled numerical features where required

– Ensured feature consistency across training and prediction phases

**Tools Used:**

– pandas, numpy

– sklearn.preprocessing

## 4. Exploratory Data Analysis (EDA):-

**Insights Generated:**

– Clear separation between species using petal measurements

– Setosa species is easily distinguishable from others

– Versicolor and Virginica show partial overlap

– Petal length and petal width are the most influential features

**Visualizations:**

– Bar charts for feature comparison

– Line plots to observe feature trends

– Pie chart for species distribution

## 5. Machine Learning Models Used:-

| Model | Accuracy (%) | Remarks |
|---|---|---|
| Logistic Regression | 96 | Strong baseline model |
| K-Nearest Neighbors (KNN) | 97 | Sensitive to feature scaling |
| Decision Tree | 95 | Captures non-linear patterns |
| Support Vector Machine (SVM) | 96 | Effective boundary separation |
| **Random Forest** | **98** | Best overall performance |

**Selected Model:**

- Random Forest Classifier (Accuracy: ~98%)

**Reason:**

Random Forest provided the highest accuracy and robust performance by combining multiple decision trees and reducing overfitting.

## 6. Model Evaluation:-

– Accuracy Score

– Precision, Recall, and F1-score

– Confusion Matrix for class-wise performance

The trained model successfully classified all three iris species with high confidence and minimal misclassification.

## 7. Tools & Technologies Used:-

| Category | Tools |
|---|---|
| Programming | Python |
| Data Analysis | Pandas, NumPy |
| Visualization | Matplotlib, Seaborn |
| Machine Learning | Scikit-learn |
| Environment | Google Colab |
| Model Saving | Pickle |

## 8. Final Output & Business Value:-

**Outcomes:**

– Successfully trained and evaluated multiple ML classification models

– Identified the best-performing algorithm

– Built a reusable ML pipeline

– Saved trained model for future predictions

**Business / Practical Impact:**

– Enables automated flower species classification

– Demonstrates real-world application of ML classification

– Can be extended to other biological or classification problems

## 9. Solution:-

This project delivers a Machine Learning-based classification system that accurately predicts the species of an iris flower using its physical measurements. By leveraging supervised learning algorithms, the solution eliminates manual classification, improves accuracy, and provides a scalable approach for similar classification tasks in real-world applications.

### 10. Author:-

Khushi Bandpatte

Engineering Student | Aspiring Data Scientist & ML Engineer