

# Project Report: Iris Flower Species Classification using Machine Learning

## 1. Introduction

This project involves building a **machine learning classification model** to accurately predict the **species of iris flowers** based on four features:

- **Sepal Length**
- **Sepal Width**
- **Petal Length**
- **Petal Width**

The three iris species being classified are:

- *Iris-setosa*
- *Iris-versicolor*
- *Iris-virginica*

We use the well-known **Iris dataset**, which contains 150 records, to train and evaluate the model.

---

## 2. Dataset Description

The dataset consists of **150 instances**, each describing an iris flower with the following attributes:

Feature	Description
Sepal Length	Length of the sepal (cm)
Sepal Width	Width of the sepal (cm)
Petal Length	Length of the petal (cm)
Petal Width	Width of the petal (cm)
Class Label	Species of the flower

There are **50 samples for each class**, making it a balanced dataset.

---

## 3. Data Preprocessing

Steps involved:

- Data loaded into a DataFrame
  - Features and target label separated
  - Target labels were encoded using LabelEncoder
  - The dataset was split into **training (80%)** and **testing (20%)** sets using `train_test_split`
- 

#### 4. Model Selection

We used the **Random Forest Classifier**, a powerful ensemble algorithm that combines multiple decision trees to improve classification accuracy and reduce overfitting.

Model details:

- **RandomForestClassifier** from `sklearn.ensemble`
  - `n_estimators=100` (100 trees)
  - `random_state=42` for reproducibility
- 

#### 5. Model Training and Testing

The model was trained using the training data and then evaluated on the test set.

```
model = RandomForestClassifier(n_estimators=100, random_state=42)
```

```
model.fit(X_train, y_train)
```

---

#### 6. Evaluation Metrics

We evaluated the model using:

- **Accuracy Score**
- **Confusion Matrix**
- **Classification Report**

**Accuracy Achieved:** *100%* on test data (in this run)

**Classification Report:**

Class	Precision	Recall	F1-score
Setosa	1.00	1.00	1.00

Class	Precision	Recall	F1-score
-------	-----------	--------	----------

Versicolor	1.00	1.00	1.00
------------	------	------	------

Virginica	1.00	1.00	1.00
-----------	------	------	------

### Confusion Matrix Visualization:

A heatmap showed all predictions were correctly classified.

---

## 7. Predictions on New Data

The trained model can be used to classify new iris flower data:

```
sample = [[5.1, 3.5, 1.4, 0.2]]
```

```
prediction = model.predict(sample)
```

**Output:** Predicted Species: Iris-setosa

---

## 8. Conclusion

- The Random Forest model was able to perfectly classify iris species with 100% accuracy.
  - The model demonstrates that **petal and sepal dimensions are strong indicators of species identity**.
  - The dataset's simplicity and separability helped achieve excellent performance with minimal tuning.
- 

## 9. Future Improvements

- Try other classifiers like **SVM, KNN, or Logistic Regression**
  - Use **cross-validation** to better assess model performance
  - Deploy the model via **Flask or Streamlit** for a web interface
- 

## 10. Tools and Libraries Used

- **Python 3.x**
- **Pandas**

- **Scikit-learn**
- **Matplotlib**
- **Seaborn**

---

Let me know if you'd like this as a formatted **PDF/Word file**, or if you want a **PowerPoint presentation** version of this report.