Building a Classification Model using Iris Dataset (Random Forest)

1. Iris Dataset
The Iris dataset contains 150 flower samples with 4 features: Sepal Length, Sepal Width, Petal Length, Petal Width.
Target classes: 0 = Setosa, 1 = Versicolor, 2 = Virginica.

2. Input and Output
X = iris.data → Shape (150, 4)
Y = iris.target → Shape (150,)

3. Random Forest Algorithm
Random Forest is an ensemble of many decision trees. Each tree is trained on random samples and random features. Final prediction is based on majority voting.

4. Training the Model
clf.fit(X, Y) trains the forest on all data samples. Each tree learns different decision boundaries.

5. Feature Importance
Petal length and petal width are the most important features. This means the model mostly uses them for classification.

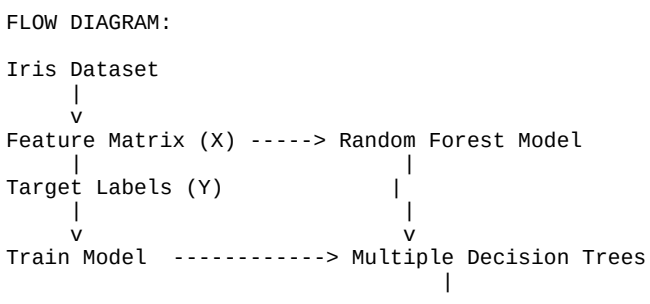
6. Prediction
Model predicts class for new input values. predict_proba gives probability for each class.

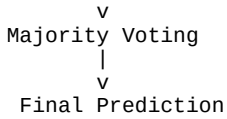
7. Train-Test Split
80% data used for training, 20% for testing. Model is evaluated on unseen data.

8. Accuracy Calculation
Accuracy = Correct Predictions / Total Predictions Your model achieved 93.33% accuracy.

9. Conclusion
Random Forest gives high accuracy and avoids overfitting. It is widely used in real-world classification problems.

10. Model Flow Diagram




v
Majority Voting
|
v
Final Prediction