Iris Dataset — Decision Tree Classifier Report

Date: August 25, 2025

1. Overview

This report presents a Decision Tree classifier trained on the Iris dataset. The goal is to classify flowers into three species: Setosa, Versicolor, and Virginica, based on four numeric features: sepal_length, sepal_width, petal_length, and petal_width.

2. Dataset Summary

Total samples: 150 | Features: 4 | Classes: 3

Class Mapping

Encoded	Label
0	setosa
1	versicolor
2	virginica

3. Preprocessing & Split

• Encoded the target label (species) using Label Encoding.
• Used raw numeric features without scaling (Decision Trees are scale-invariant).
• Train/Test split: 80% train, 20% test with stratification by class.

4. Model Configuration

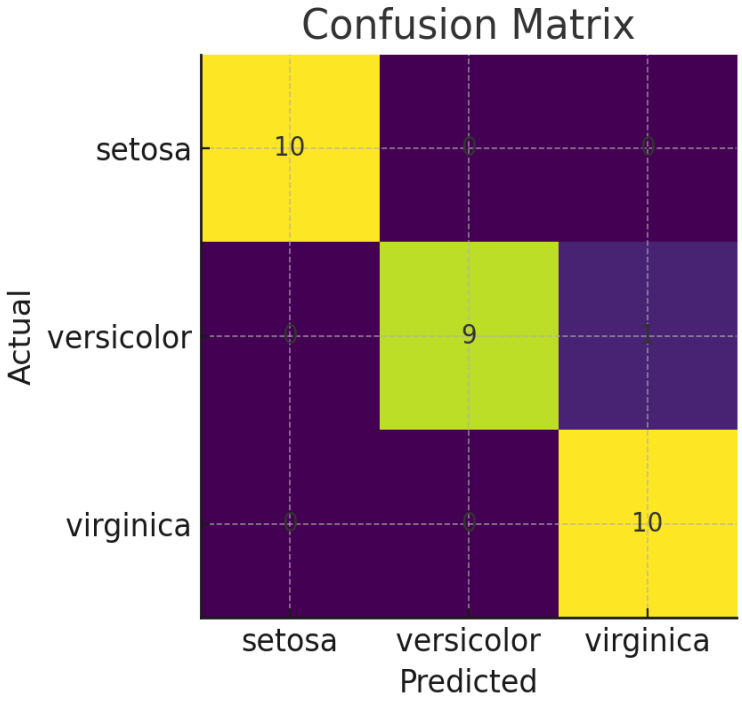
Algorithm: DecisionTreeClassifier
Parameters: max_depth=3, random_state=42

5. Evaluation Metrics

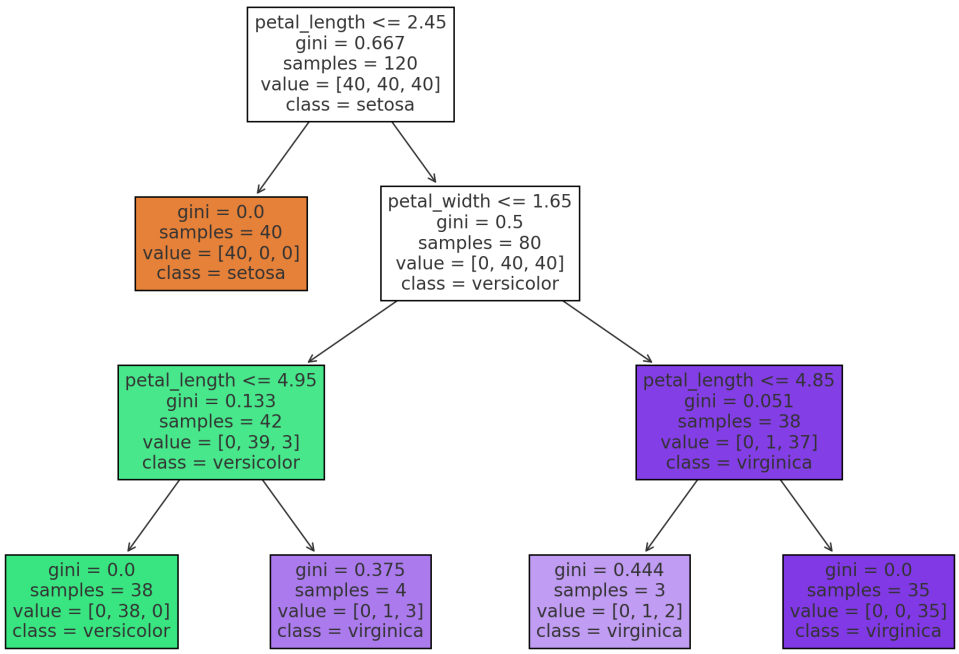
Overall Accuracy: **0.9667**

Class	Precision	Recall	F1-Score	Support
setosa	1.00	1.00	1.00	10
versicolor	1.00	0.90	0.95	10
virginica	0.91	1.00	0.95	10
macro avg	0.97	0.97	0.97	30
weighted avg	0.97	0.97	0.97	30

6. Confusion Matrix



7. Decision Tree Visualization



8. Notes & Next Steps

• The model achieves strong performance on the Iris dataset. Further improvements could include cross-validation for more robust estimates, grid search for hyperparameter tuning (e.g., max_depth,

min_samples_split), and comparing with other algorithms like Random Forest or SVM.