Module 5: Critical Thinking

Random Forest Classifier

Ryan Thompson

Colorado State University - Global

CSC 580

Dr. Farr

4 July 2025

Iris Random Forest Classifier

The Random Forest classifier implemented on the Iris dataset demonstrates a robust approach for classifying flower species using ensemble methods within machine learning. The classifier was trained on four numerical features from the Iris dataset: sepal length, sepal width, petal length, and petal width, which are well known predictors for classifying setosa, versicolor, and virginica species.

The data preparation involved splitting the dataset randomly into training (75%) and test (25%) subsets to evaluate the model’s performance objectively. The categorical target variable representing species was factored into numerical labels, a standard preprocessing step that facilitates training with scikit-learn models. By employing the RandomForestClassifier from scikit-learn with multiple estimators, the classifier utilizes the ensemble of decision trees to reduce variance and improve generalization.

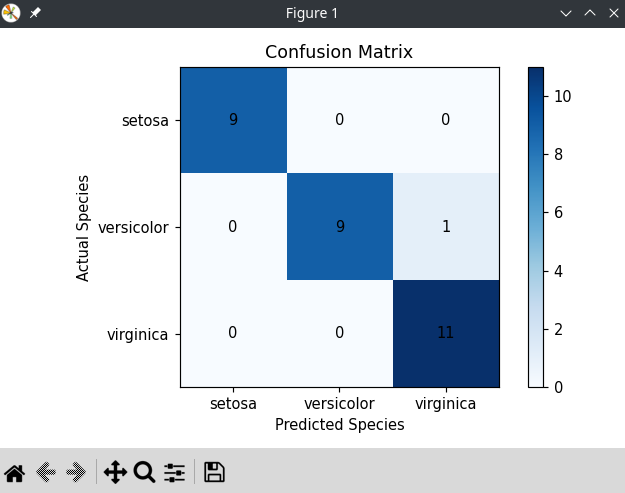
Model training involved fitting the Random Forest on the training data using the four feature columns, leveraging parallel computation to speed up training. After training, predictions were generated on the test set, which were then evaluated using a confusion matrix. The matrix revealed how accurately the model predicted each species and identified areas where misclassifications occurred. For visualization, a heatmap of the confusion matrix was plotted, providing an immediate overview of model performance across the classes.

Feature importance was extracted from the trained model, showing the relative significance of each feature in determining the species classification. The model revealed that petal length and petal width were the most important features, aligning with the known separability of Iris species based on these characteristics. This insight can guide future data collection efforts by emphasizing the most predictive features for similar classification tasks.

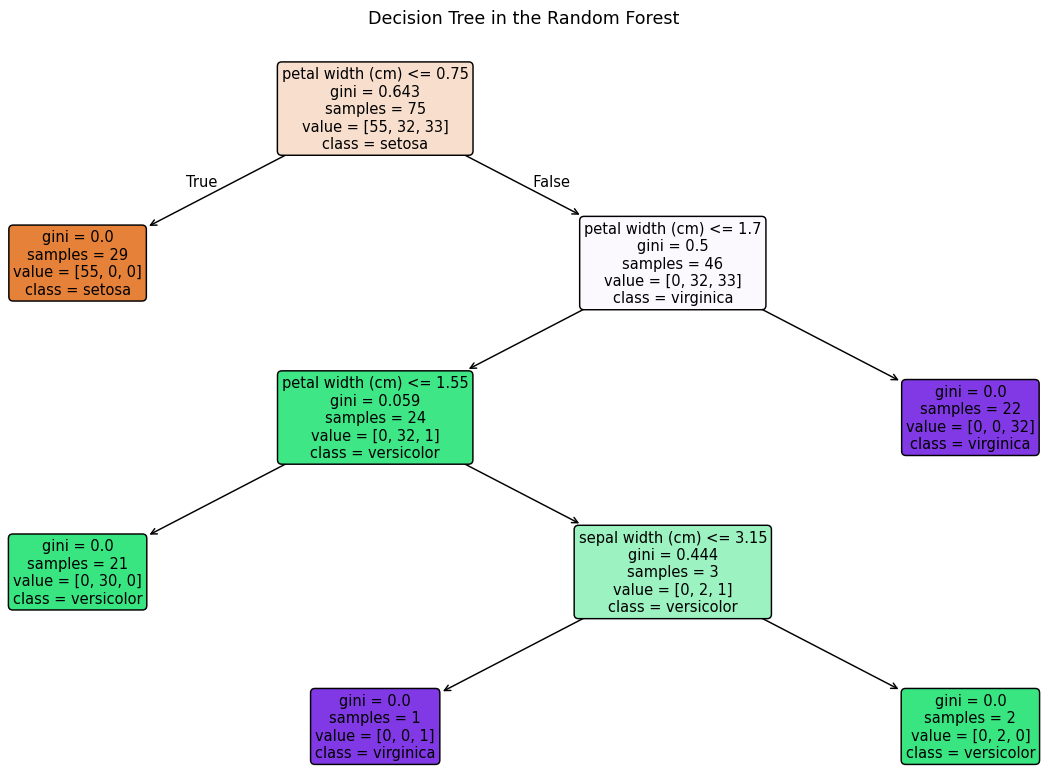
Additionally, the code included a visualization of a single decision tree within the Random Forest, offering interpretability into how individual trees contribute to the ensemble's decision making process. This is particularly beneficial for understanding the structure of the classifier and for educational purposes in illustrating the splits that lead to class predictions.

Overall, the Random Forest classifier achieved high accuracy in classifying the Iris species, demonstrating the effectiveness of ensemble learning on structured datasets with clear feature target relationships. The model is also robust against overfitting due to its bagging strategy, and its ability to provide feature importances makes it a valuable tool for both prediction and interpretation within supervised machine learning tasks.

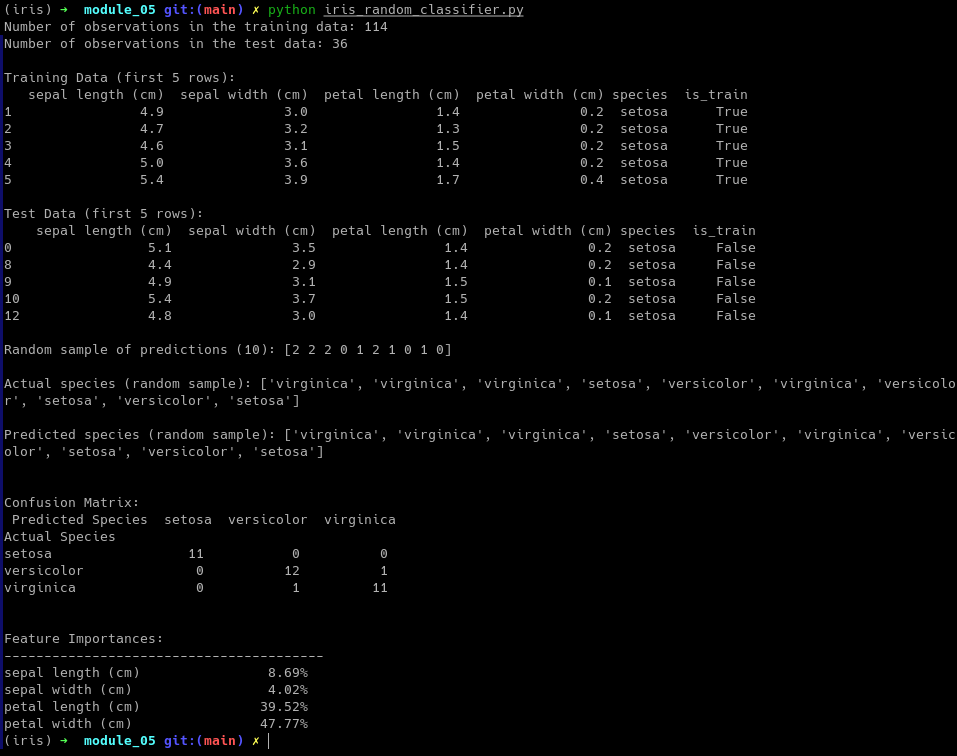
Program Outputs



*Confusion Matrix*



*Decision Tree Output*



*Terminal Output*

References

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825-2830.