Experiment 5

Classification Using J48

Aim: To demonstrate Classification process on iris.arff dataset using j48 algorithm with percentage split.

Tasks:

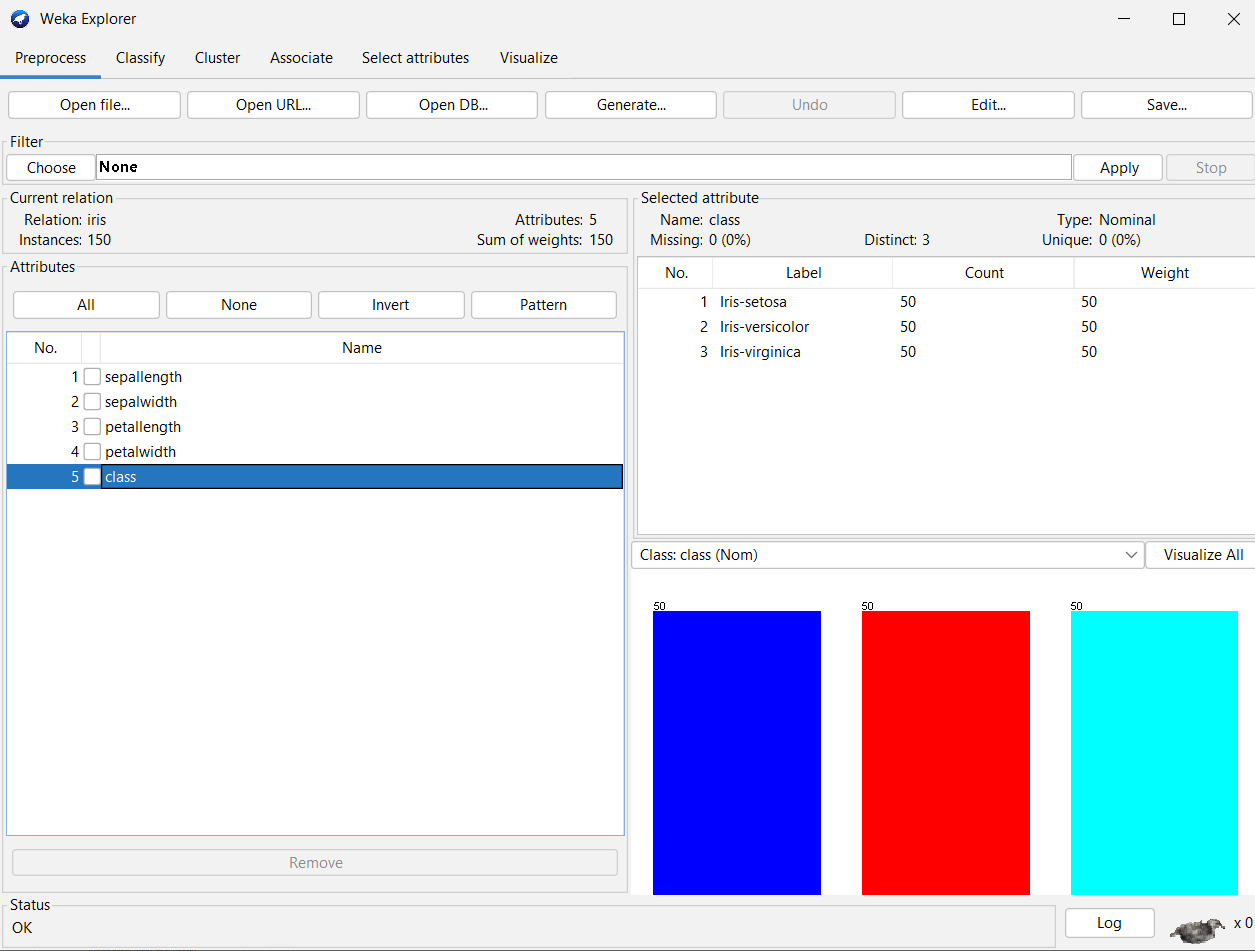
1. Load iris.arff dataset and explore it.
2. Build a classification model using J48 algorithm with percentage split.
3. Make predictions on new data.

Task 1: Load iris.arff dataset and explore it.

Load iris.arff from the Weka’s data folder.

Observations:

|  |  |  |
| --- | --- | --- |
| **Attribute** | **Type** | **Range / States** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| Number of Instances: | | |



Task 2: Build a classification model using J48 algorithm with percentage split.

Classification is a process of determining the class (state) of the given instance. Examples:

Determining Play or Not play based on weather conditions.

Determining the digit (0 – 9) given the image pixel data.

Determining the Spam or Not-spam based on mail text.

J48 is Weka’s Java implementation of the C4.5 algorithm. It can generate pruned or unpruned tress with both nominal and numerical attributes for classification.

Percentage Split:

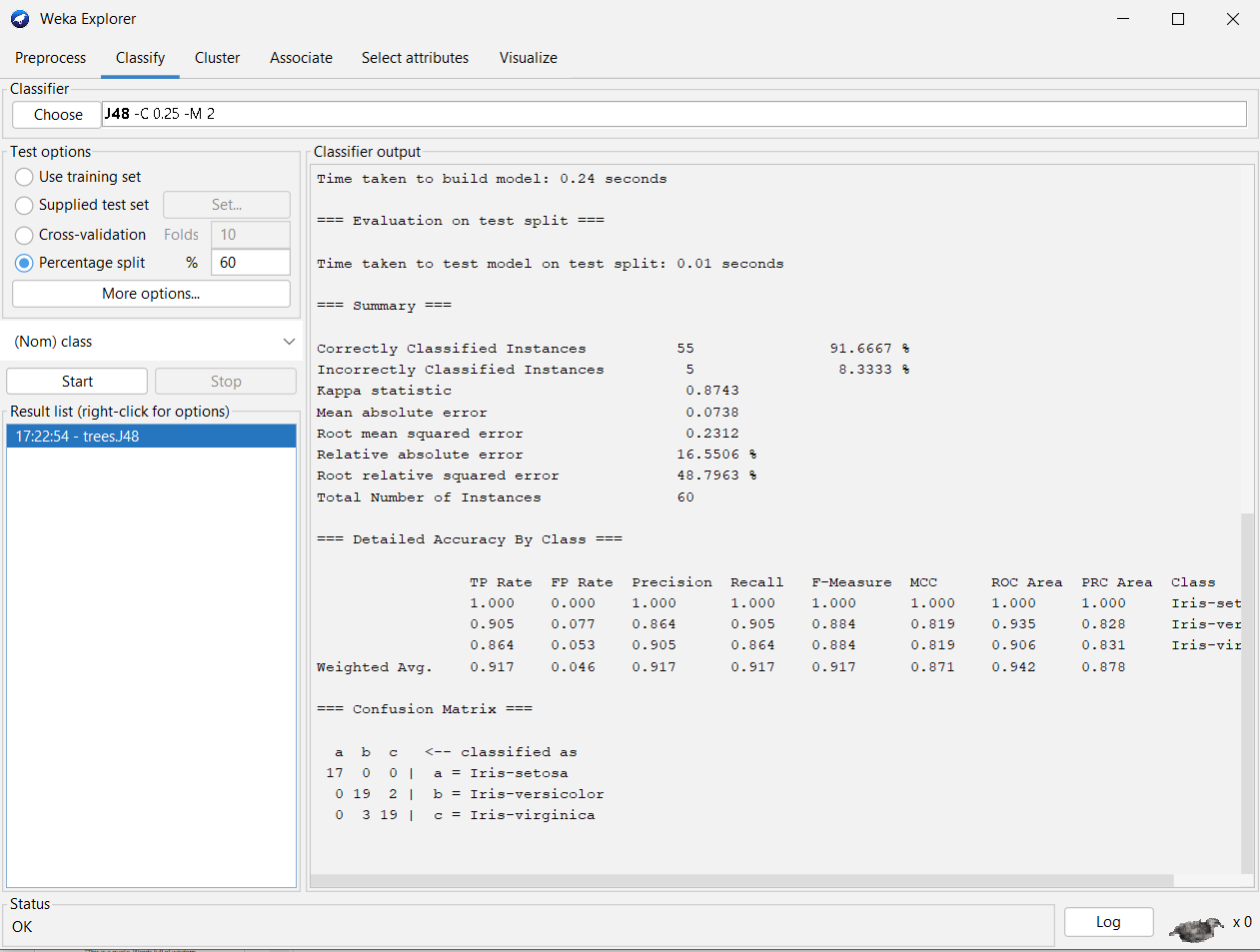
Splits the data into training and test subsets.

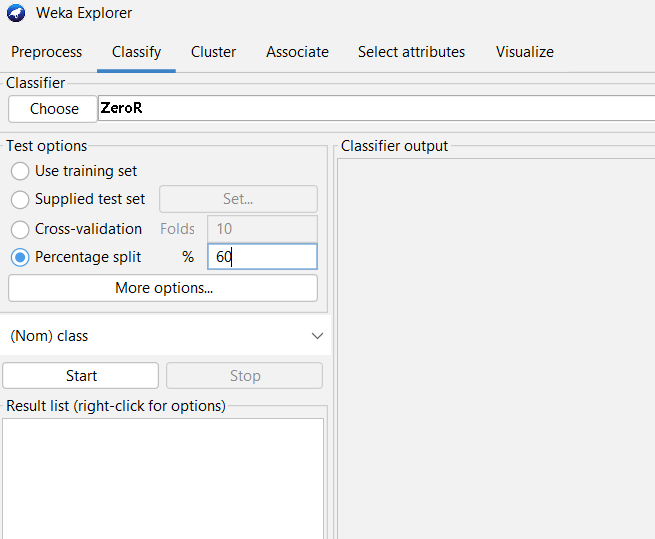
Training set is used to build the model.

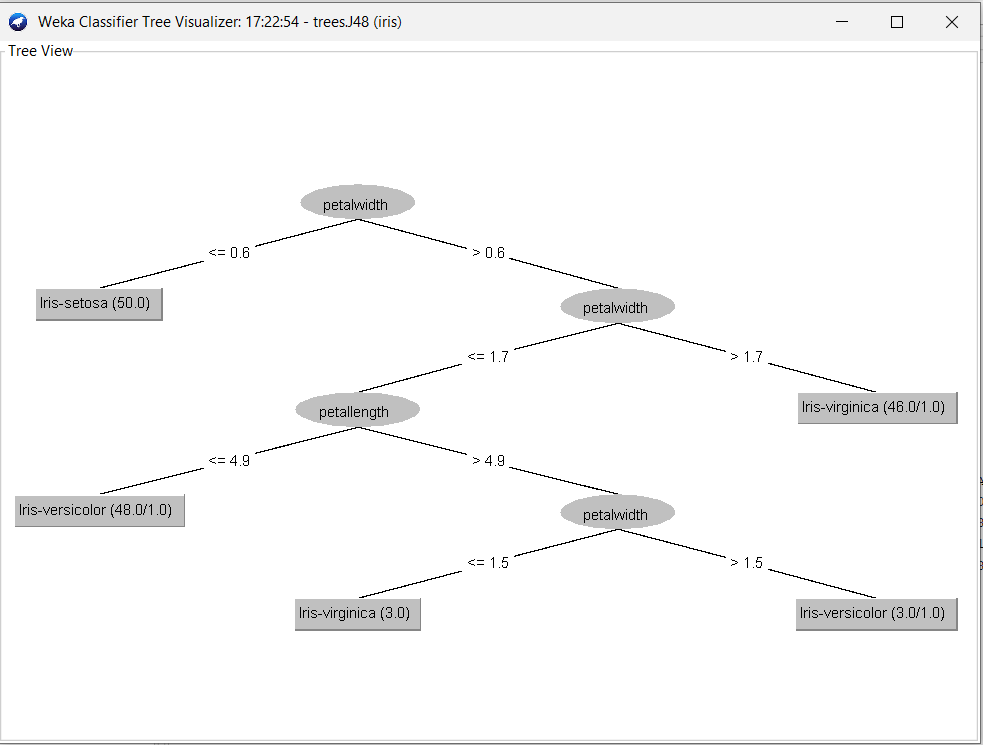
Test set is used to evaluate the model performance.

Steps to build the model:

* 1. Click on Classify and select Percentage split with required training percentage under Test options group.
  2. Select Choose 🡪 classifiers 🡪 trees 🡪 J48.
  3. Clock on Start.
  4. Right click on the model and click on Visualize tree







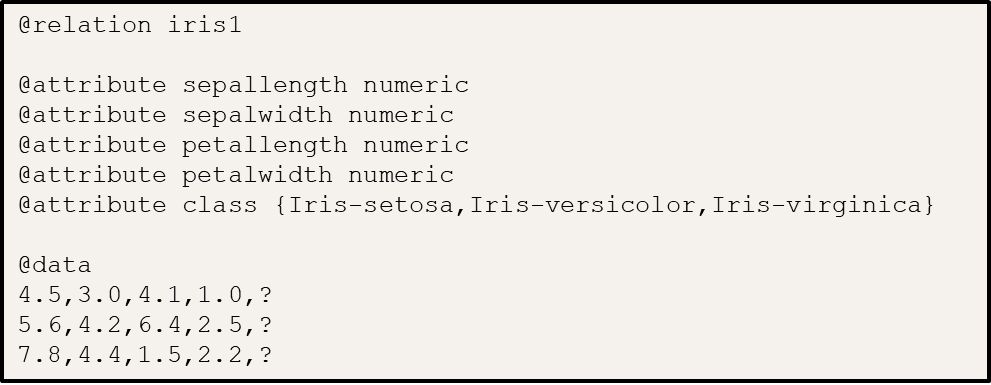
Observations:

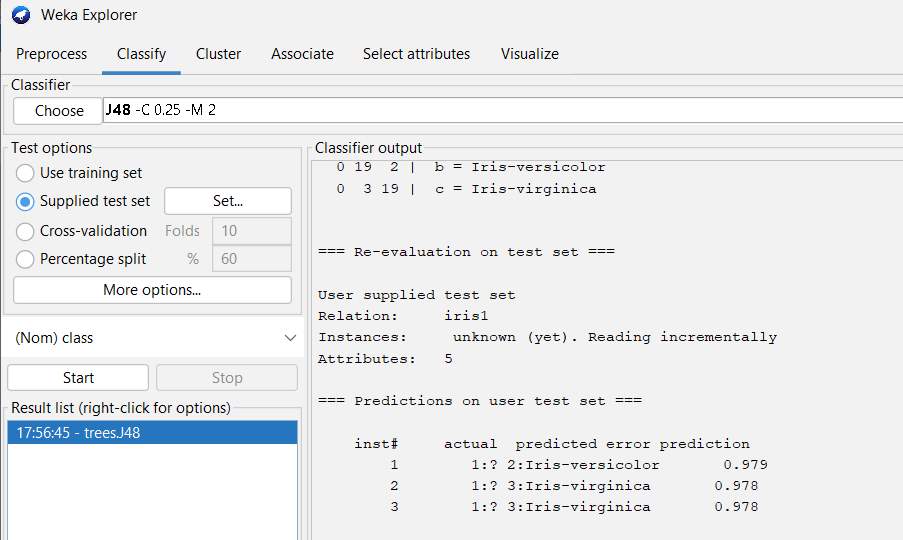
* + Total number of instances:
  + Correctly classified instances:
  + Incorrectly classified instances:
  + Accuracy:
  + Calculation of Accuracy from Confusion Matrix:

Task 3: Make predictions on new data

Steps:

* 1. Create an ARFF file with unlabeled (use ? in the place of class label) instances.
  2. On the “Classify” tab, select the “Supplied test set” option in the “Test options” pane.
  3. Click the “Set” button, click the “Open file” button on the options window and select the new dataset.
  4. Click the “More options…” button and for the “Output predictions” option click the “Choose” button and select “PlainText”.
  5. Right click on the model in the “Results list” pane and Select “Re-evaluate model on current test set”.





Observations;

|  |  |
| --- | --- |
| Instance No. | Predicted class |
|  |  |

Conclusion: