Experiment 6

Classification Using J48

Aim: To demonstrate Classification process on StudentResult.arff dataset using j48 algorithm with cross-validation

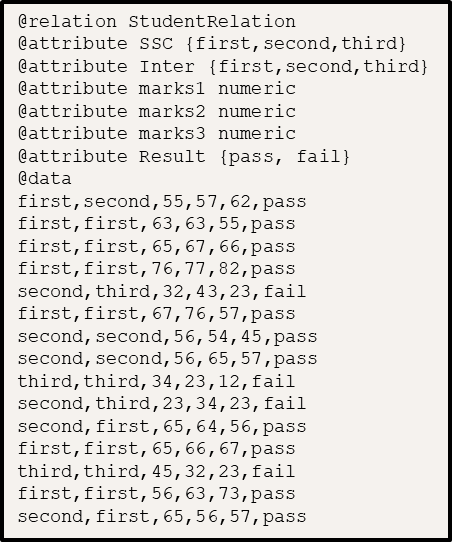
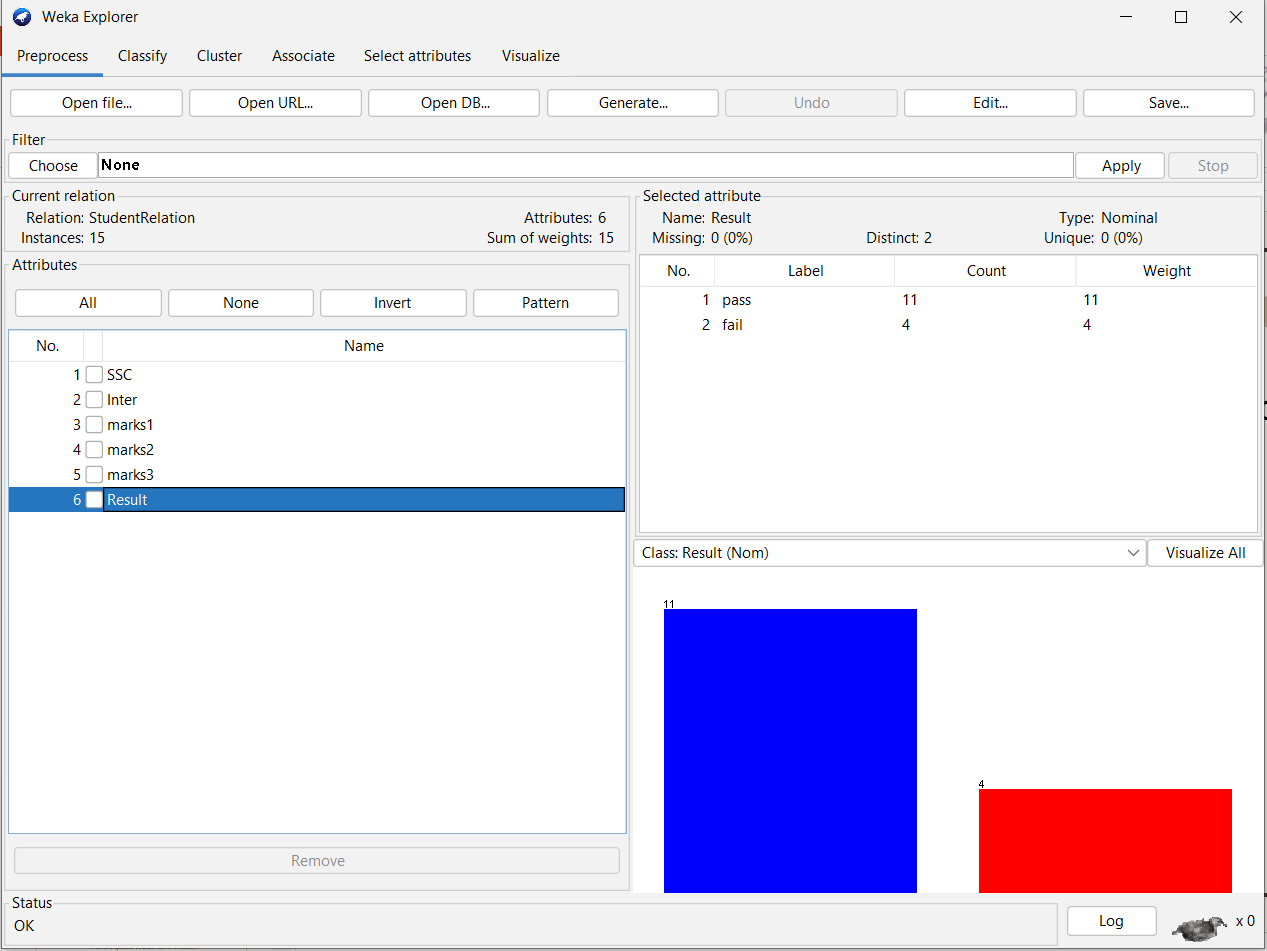
Tasks:

1. Create StudentResult.arff dataset and load it into Weka.
2. Build a classification model using J48 algorithm with k-fold cross validation.
3. Make predictions on new data.

Task 1: Create StudentResult.arff dataset and load it into Weka.

Create StudentResults.arff with following attributes and load it into Weka

|  |  |
| --- | --- |
| Attribute | Type |
| SSC | Nominal with states First, Second, Third. |
| Inter | Nominal with states First, Second, Third. |
| Marks1 | Numeric |
| Marks2 | Numeric |
| Marks3 | Numeric |
| Result | Nominal with states Pass & Fail. |

Task 2: Build a classification model using J48 algorithm with k-fold cross validation

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.

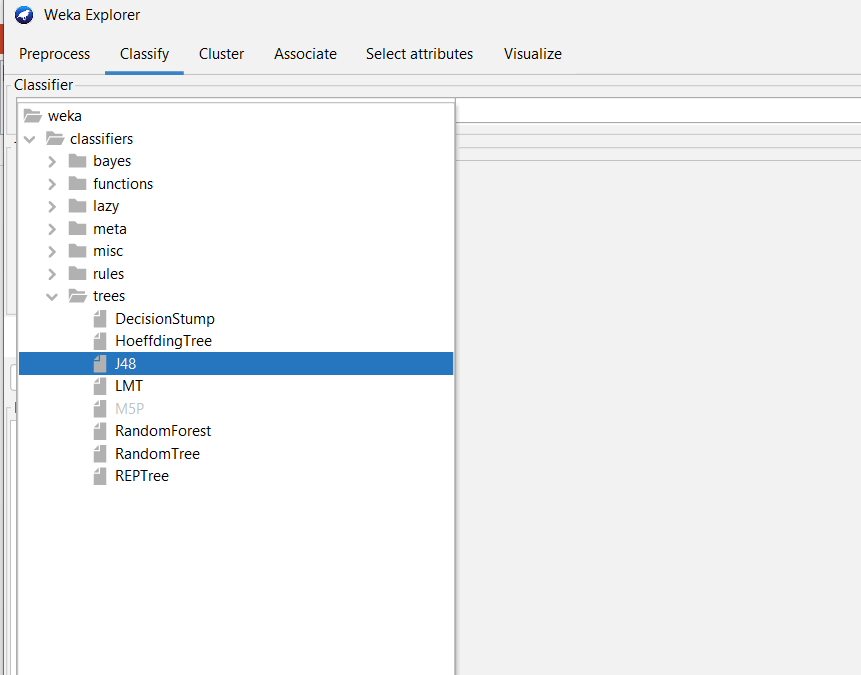
J48 is Weka’s Java implementation of the C4.5 algorithm. It can generate pruned or unpruned tress with both nominal and numerical attributes for classification.

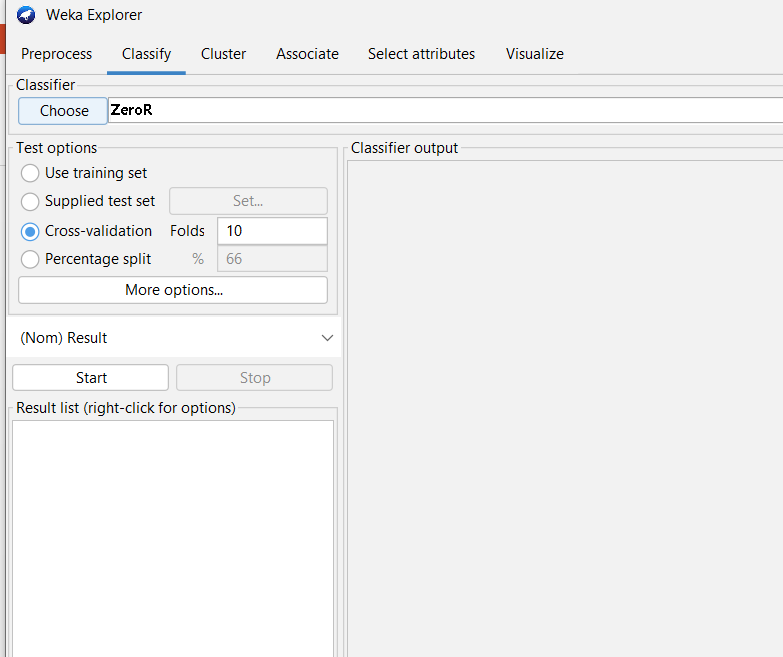
K-fold Cross Validation is a resampling procedure used to evaluate data mining models on a limited data set. It’s process is

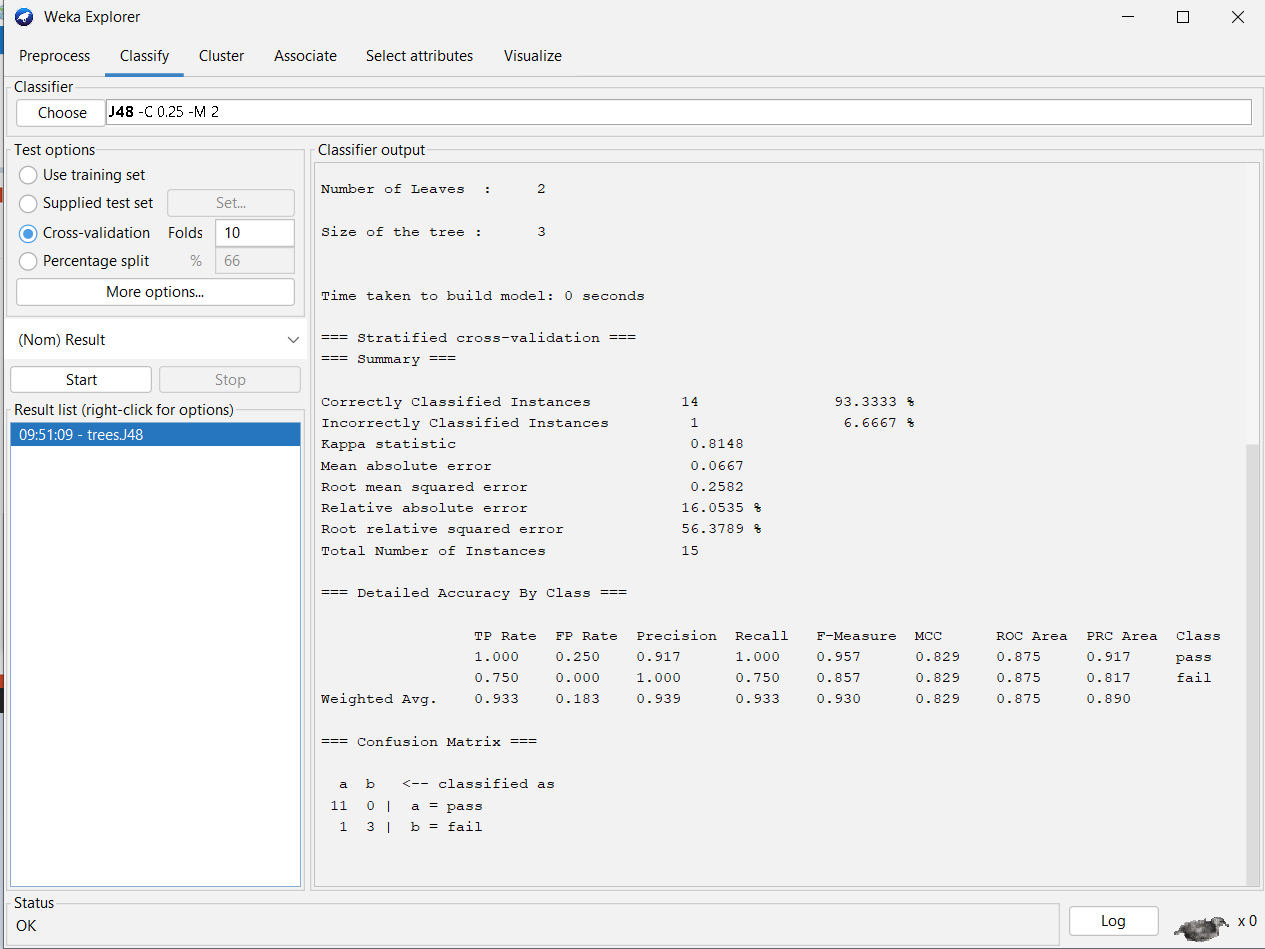
* + 1. Split the input dataset into K groups
    2. For i from 1 to k
       - Take ith group as test dataset.
       - Use remaining K-1 groups as training dataset.
       - Fit the model using training set and evaluate its performance on test set.

Steps to build the model:

* 1. Click on Classify and select Cross-validation with default 10 folds 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







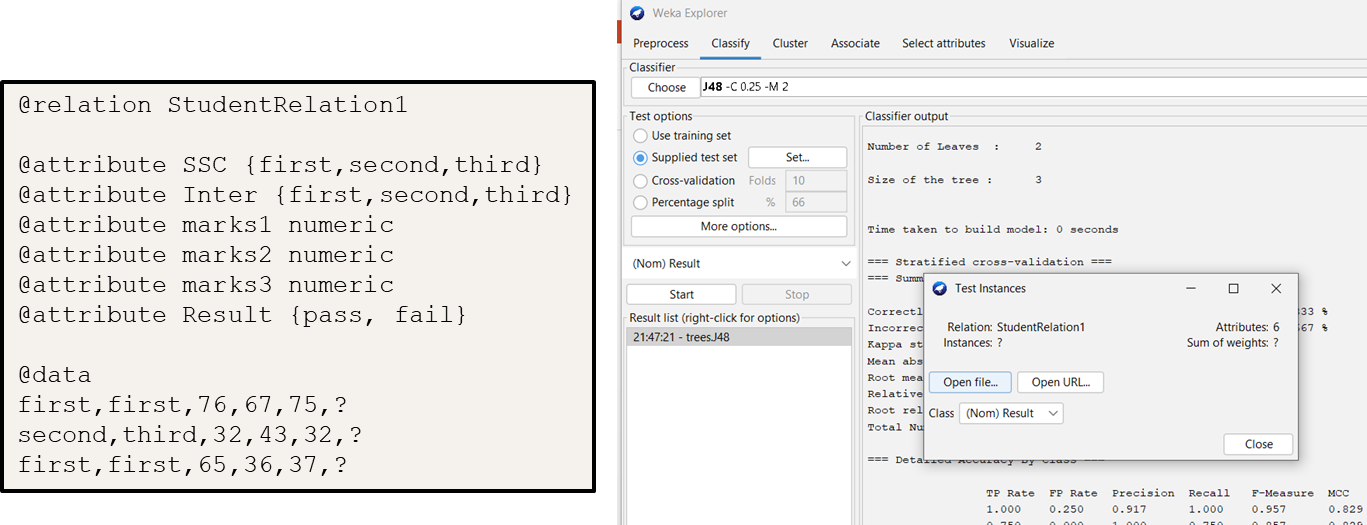
Observations:

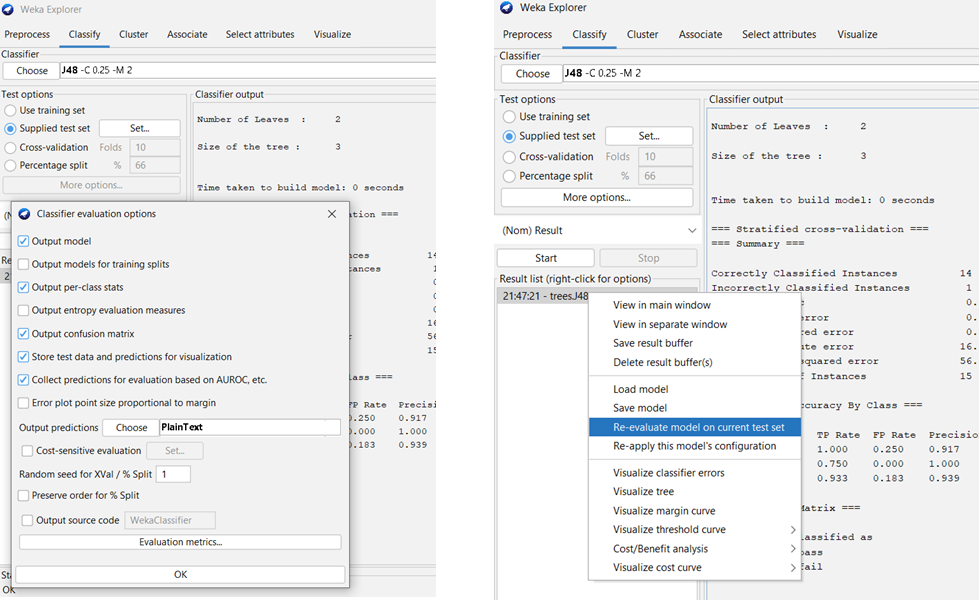
* + Total number of instances:
  + Correctly classified instances:
  + Incorrectly classified instances:
  + Accuracy:
  + Calculation of Accuracy from Confusion Matrix:
  + Tree generated by J48 algorithm for the given dataset is:

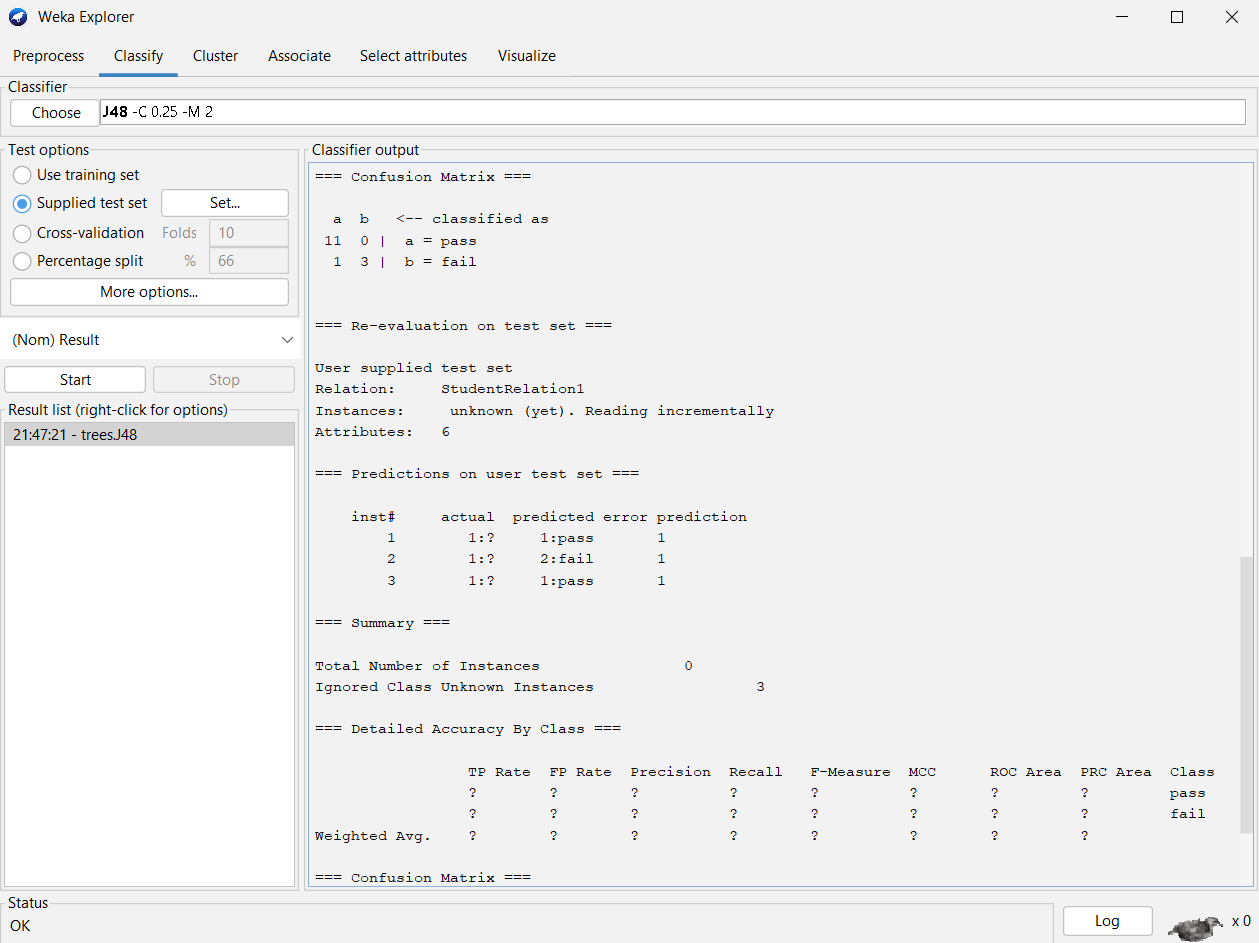
Task 3: Make predictions on new data

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: