Sub : DWM Year/ Semester : TE/V

Experiment Number: 07

Aim: Implementation of Decision tree classifier using WEKA

Problem Statement: To implement Decision Tree in weka Using J48 algorithm.

Objective:

The objective of the learning Decision tree classifier algorithm is to build prdictive model that accurately predict the class labels of previously unkonw records.

Theory: A decision tree is a tree in which each branch node represents a choice between a number of alternatives, and each leaf node represents a decision.

Decision tree are commonly used for gaining information for the purpose of decision - making. Decision tree starts with a root node on which it is for users to take actions. From this node, users split each node recursively according to decision tree learning algorithm. The final result is a decision tree in which each branch represents a possible scenario of decision and its outcome.

ID3 is a simple decision tree learning algorithm developed by Ross Quinlan (1983). The basic idea of ID3 algorithm is to construct the decision tree by employing a top-down, greedy search through the given sets to test each attribute at every tree node. In order to selectthe attribute that is most useful for classifying a given sets, we introduce a metric--- informationgain.

To find an optimal way to classify a learning set, what we need to do is to minimize the questions asked (i.e. minimizing the depth of the tree). Thus, we need some function which can measure which questions provide the most balanced splitting. The information gain metric is such a function.

Procedure:

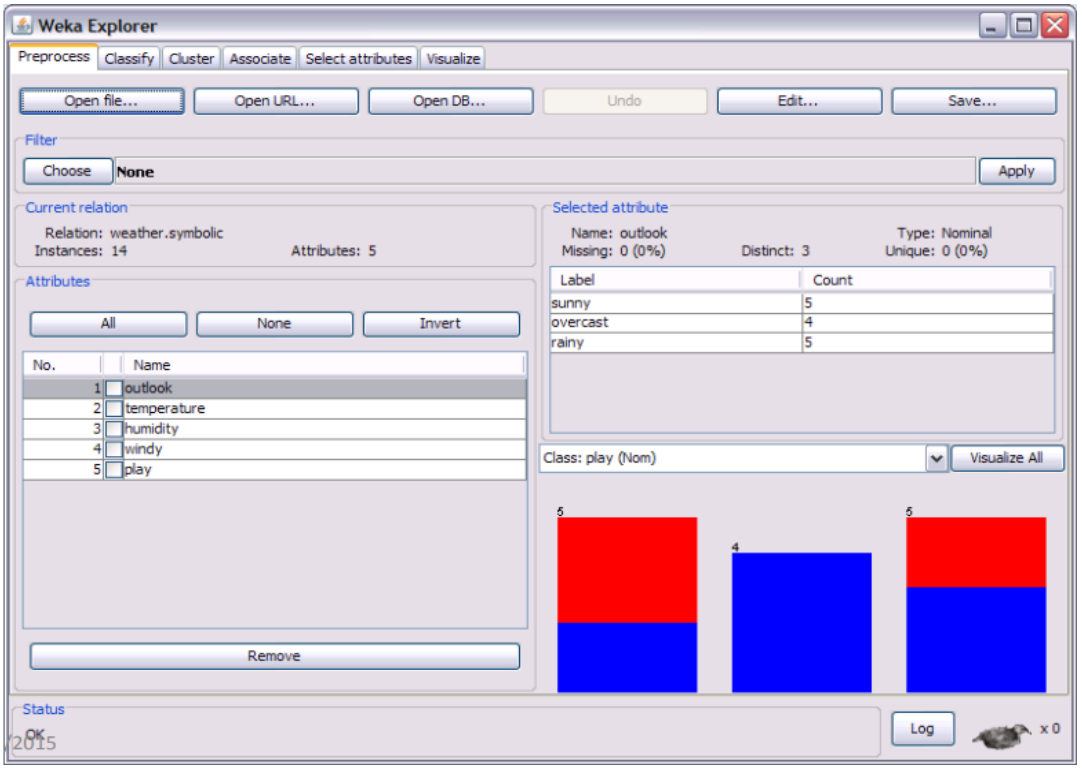
Open a data file

In “Preprocess” panel, click “Open file” button, and choose an ARFF file from “data” folder. Now the other tabs are active.

If you specify a “CSV” file, it will be automatically converted into ARFF file.

Select for example weather.nominal.arff .

Pre-process Panel



Pre-process Panel

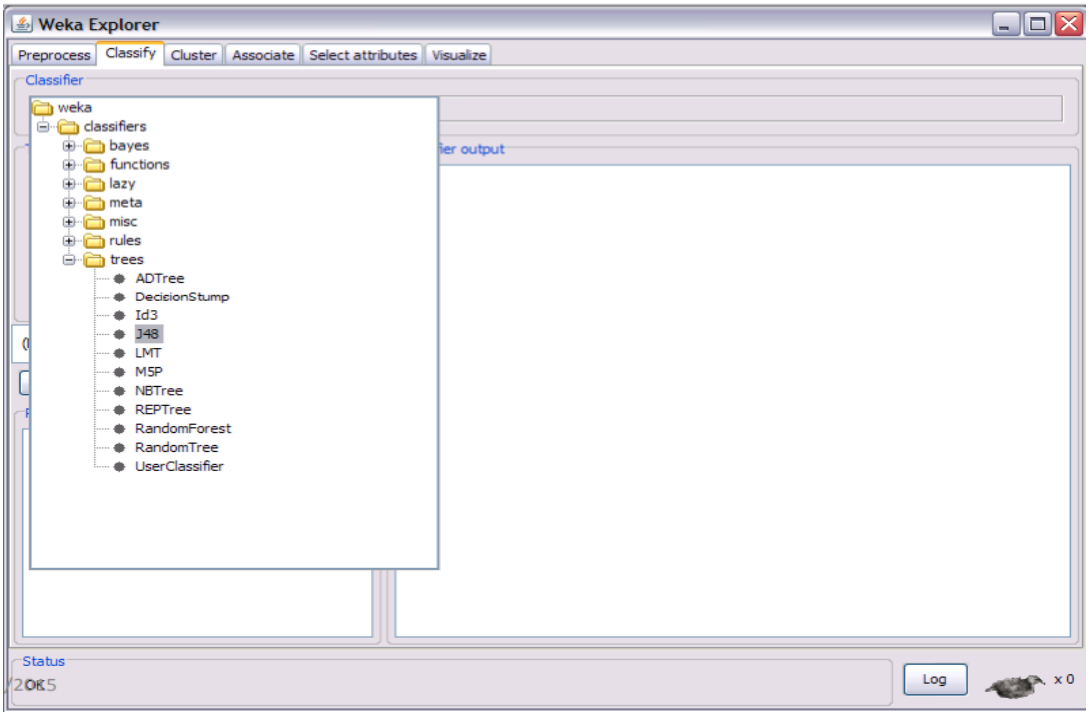
- Current relation pane: specification of data file
- Attributes and Remove button to possibly remove some attributes from the experiment.
- Statistics about selected attribute (highlighted attribute.) – If you select a numeric attribute, If you select a numeric attribute, Min, Max, Mean, and standard deviation are shown.
- Histogram shows the distribution of the class as the function of the selected attribute.
- Edit button to edit input data on a separate window – In the Edit window, right click on the caption of column or values opens a list of various available editions on data. – Undo, and Save buttons have the known function

Build a Decision Tree

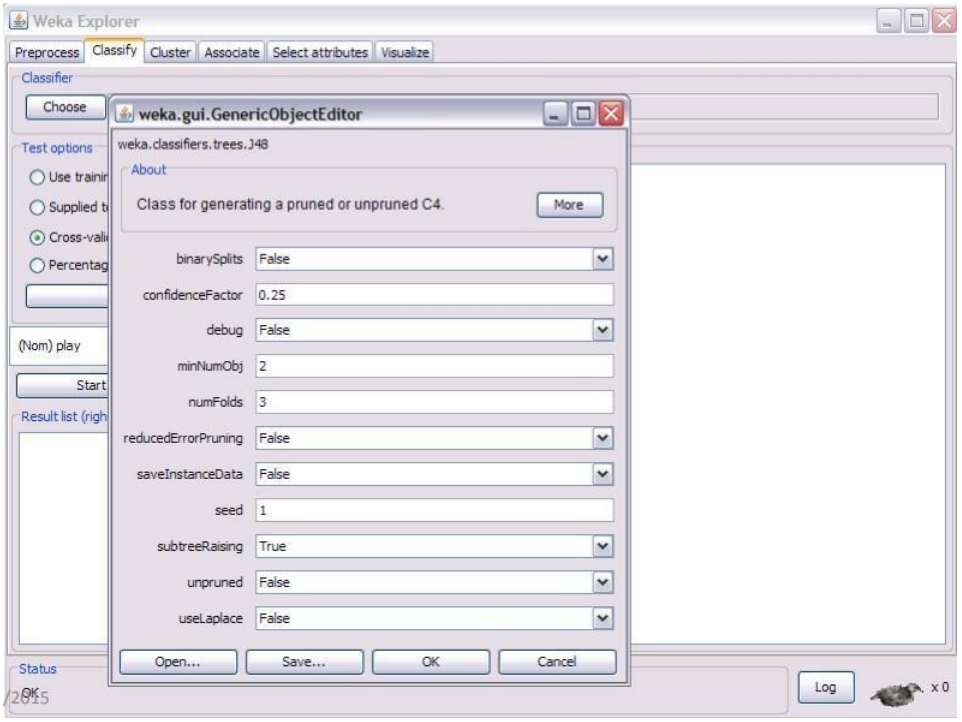
- Switch to "Classify" tab
- Select "J48" algorithm (an implementation of C4.5) by – Clicking "Choose" button – Selecting classifiers >> trees >> J48 from the Weka tree
- Invoke classifier by clicking "start" button
- Clicking the line in front of the "choose" button, opens classifier's Object Editor, in which any parameter can be changed.

- Weka keeps the results of different classifiers in the "Result List" pane.

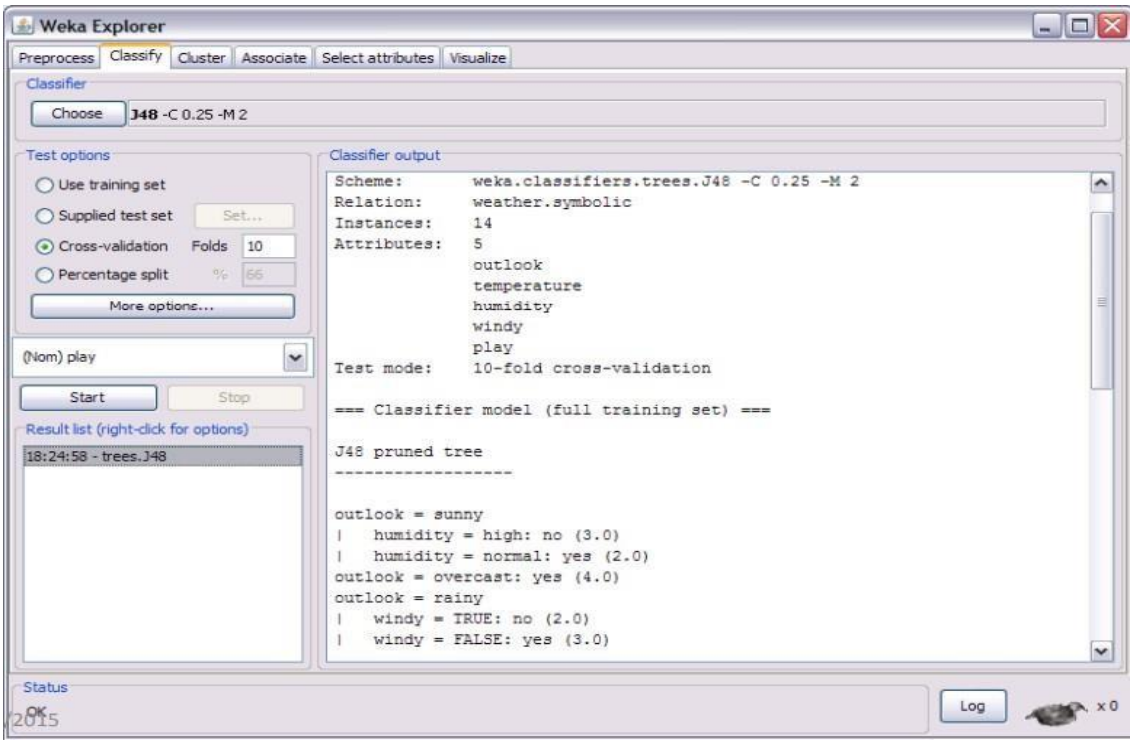
Classifier



Object Editor



Decision Tree



Classifier Output

Classifier output includes:

- Summary of the data set
- 10-fold cross validation is the default "test (evaluation) mode"
- A pruned decision tree in textual format
- A colon (:) introduces the class label assigned to the leaf, followed by number of instances reached that leaf
- Number of leaves and nodes in the decision tree
- Estimates of the tree's predictive performance
- Confusion matrix at the end
- Some other statistics

Conclusion:

Thus we have studied demonstration of decision tree classification in weka by using J48 algorithm.