Experiment No 7

Name : Dhiraj Ravindra Bodake
Roll No : 18141216

Title : Experiment with Decision Trees on Weka

Theory : Decision Trees are also known as classifn and Regresion Tree. They work by learning and to hierarchy if leading to a decision. These questions form a tree like structure and hence the names.

Each node in decision tree represents a question derived from the Feature present in yours dataset. Your dataset is split based on these questions until max. depth at the tree is reached.

- The topmost node in decision tree is root node.
- The bottom-most node is called the leaf node
- A node divided into sub-node is called Parent node and the sub-nodes are child nodes.

1. load the dataset in weka by selecting 'open file' button in weka Explorer
2. click on 'visualize All' button to view all the plots together.
3. click on 'classify' tab on the top and then click on choose button
4. From 'drop-down list' select "trees" which will open all the tree algorithms.

5. Finally select 'Rep Tree' decision tree and click on "start" button

- Decision Tree Parameters in Weka:
  - i) MaxDepth - it determines max. depth of your decision tree. By default it is which means algo. will automatically control depth.
  - ii) noPruning - Pruning means to automatically cut back on a leaf node that does not contain much info. This keeps decision tree simple and easy to interpret.
  - iii) numFolds - The specified no of folds of data will be used for purning the decision tree. The rest will be used for growing rules.
  - iv) Min Num - minimum no. of instances per leaf. If not mentioned the tree will keep splitting till all leaf nodes have only one class associated with it.

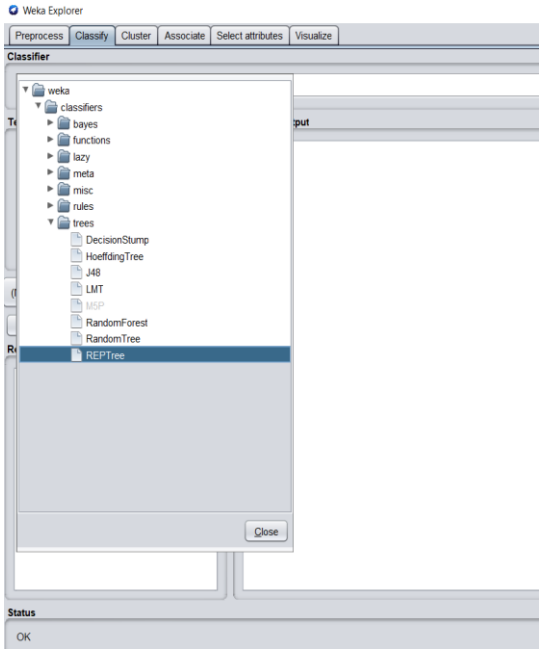
- Visualizing Decision Tree in Weka:
  1. Go to 'Result list' section and right click on your trained algo
  2. choose "Visualise tree" option and you get output.

- The values on the lines joining nodes represent splitting criteria based on values in Parent Node - Feature.

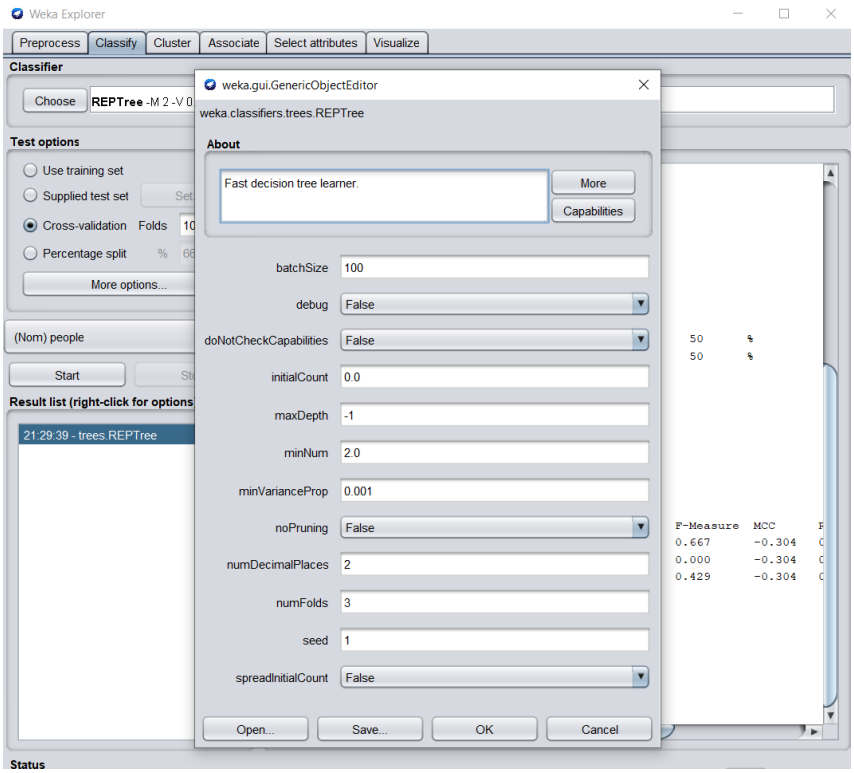
- In leaf node the value before parenthesis denotes classifh value.
- The first value in first parenthesis is the from training set in that leaf.
- Second value is the no. at instances incorrectly classified in that leaf.
- The first value in the second paranthesis is the total no. of instances from purning set in that leaf.
- The second value is the no. of instances incorrectly classified in that leaf.

OUTPUT:

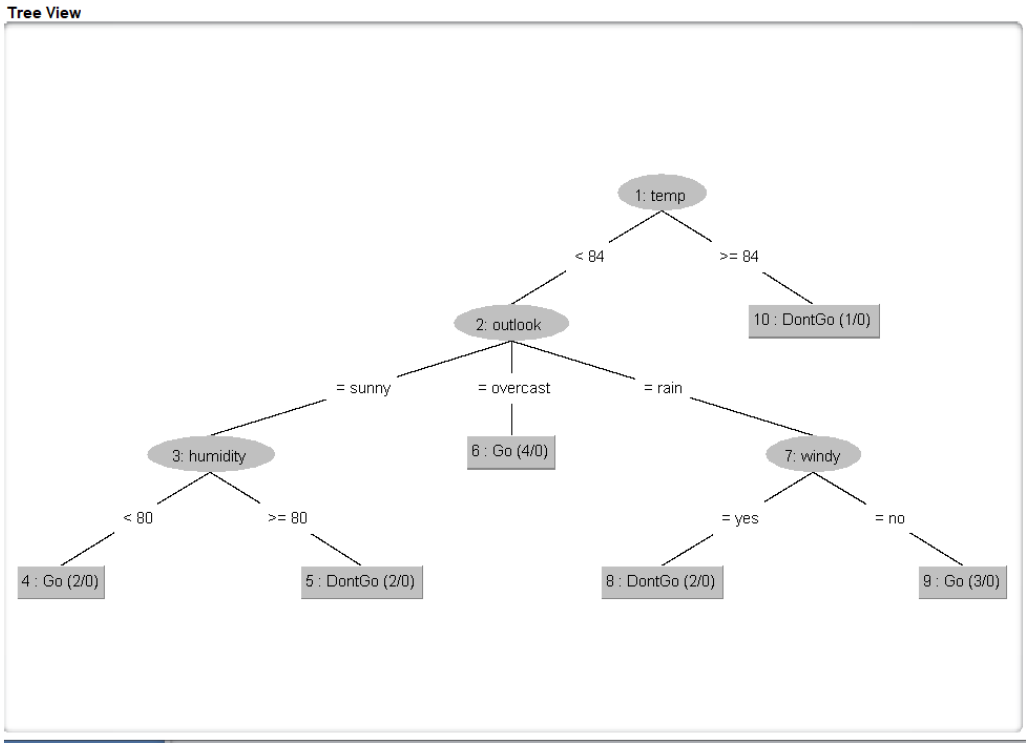
Selecting REPTree decision tree in classify window



Decision tree parameters:



Tree view of given data set



Conclusion:

Thus We studied and understood the Decision Tree on weka.