120A3051

Shreya Idate

Batch: E3

Experiment 8

Aim: Use Weka to implement the classifier-Decision Tree Algorithm

Theory:

Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.

In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.

The decisions or the test are performed on the basis of features of the given dataset.

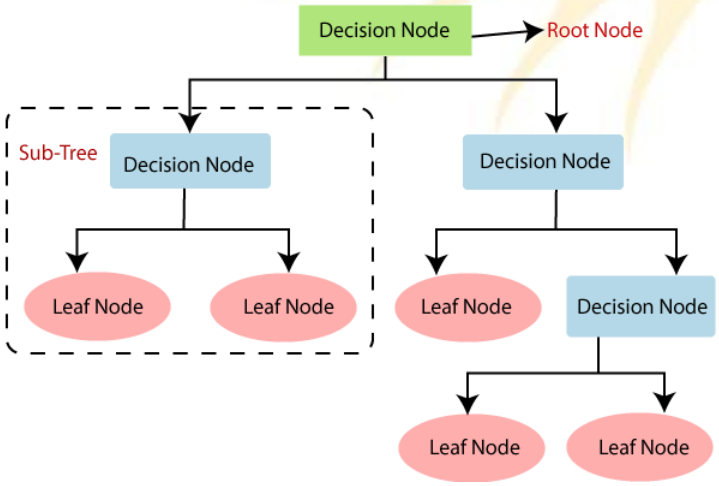
It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions.

It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure.

In order to build a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm.

A decision tree simply asks a question, and based on the answer (Yes/No), it further split the tree into subtrees.

Below diagram explains the general structure of a decision tree:



Advantages:

1. It is easy to grasp because it follows a constant method that somebody follows whereas creating any call-in real-life.
2. It is terribly helpful for the resolution of decision-related issues.
3. It helps to place confidence in all the attainable outcomes for a haul.
4. There is less demand for knowledge cleansing compared to alternative algorithms.

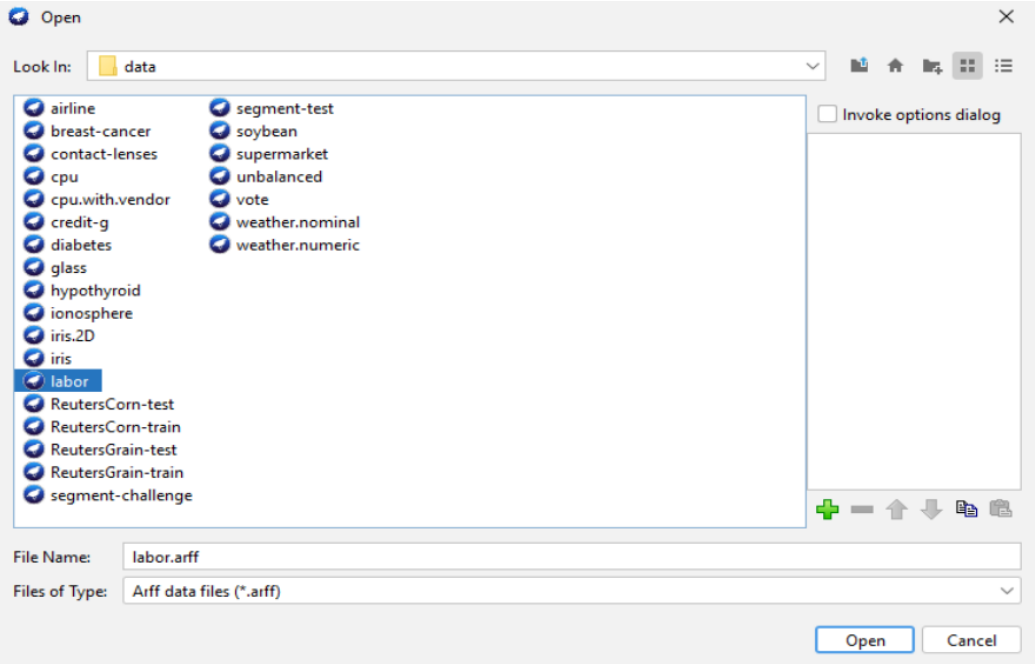
Disadvantages:

1. The decision tree contains legion layers, which makes it advanced.
2. It may have an associate overfitting issue, which might be resolved exploitation the Random Forest formula.
3. For a lot of category labels, the process quality of the choice tree could increase.

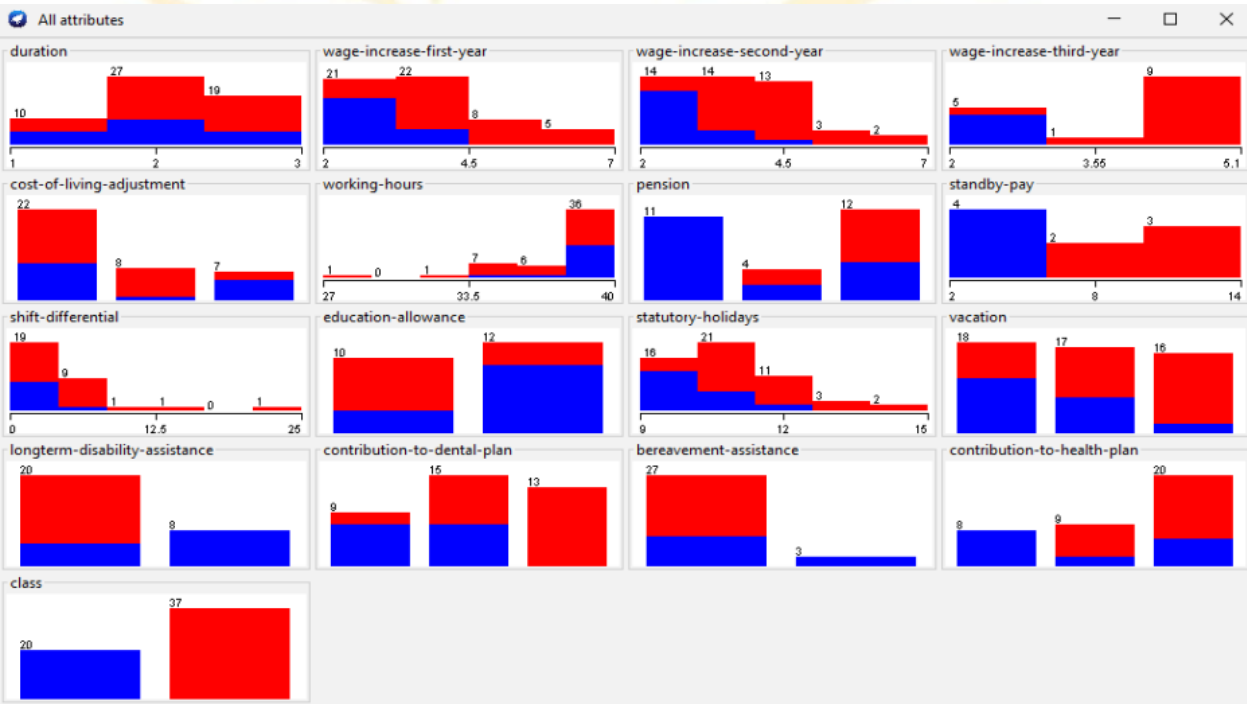
Output:

Step 1: Open Weka GUI & Select the "Explorer" option.

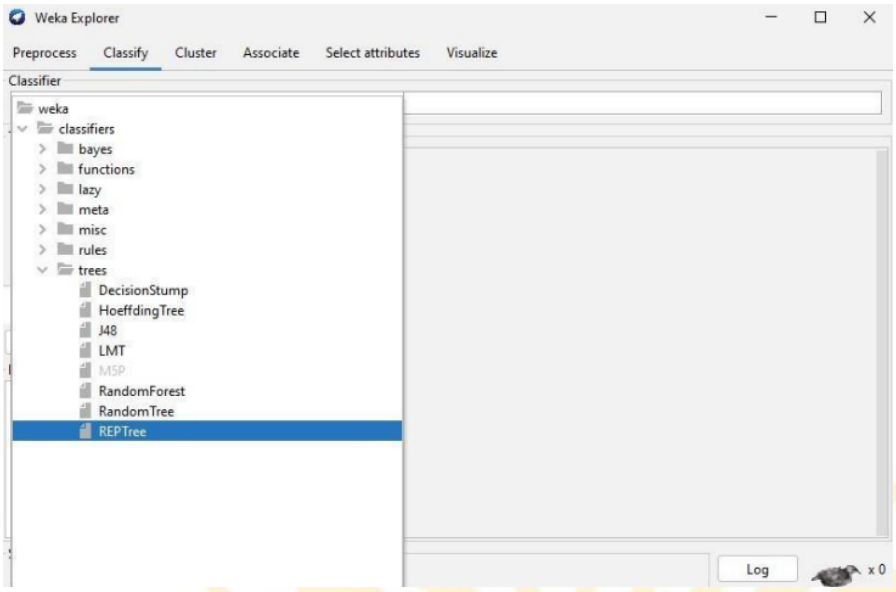
Step 2: Select “Open file” and choose your dataset.



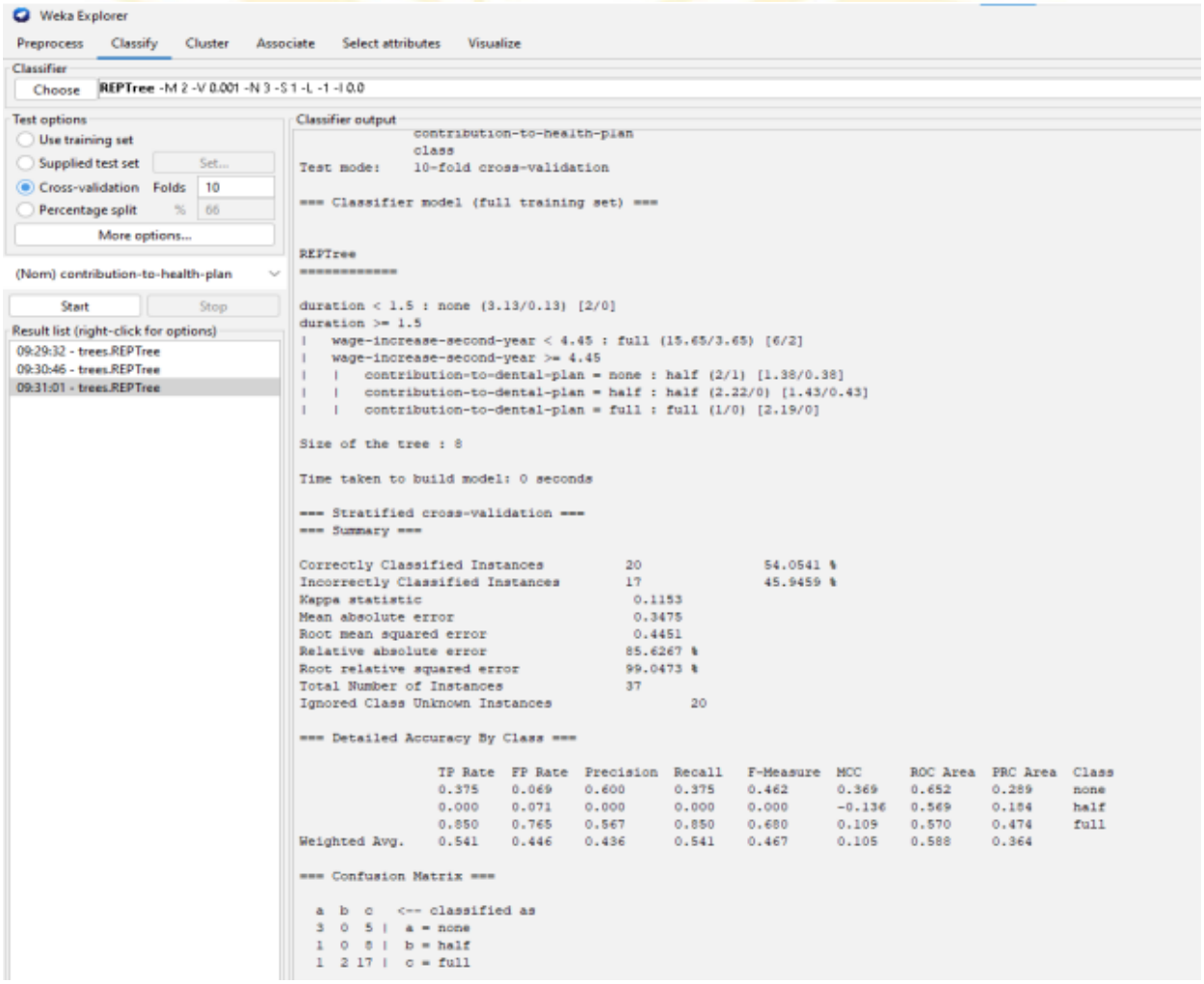
Step 3: You can even view all the plots together if you click on the “Visualize All” button



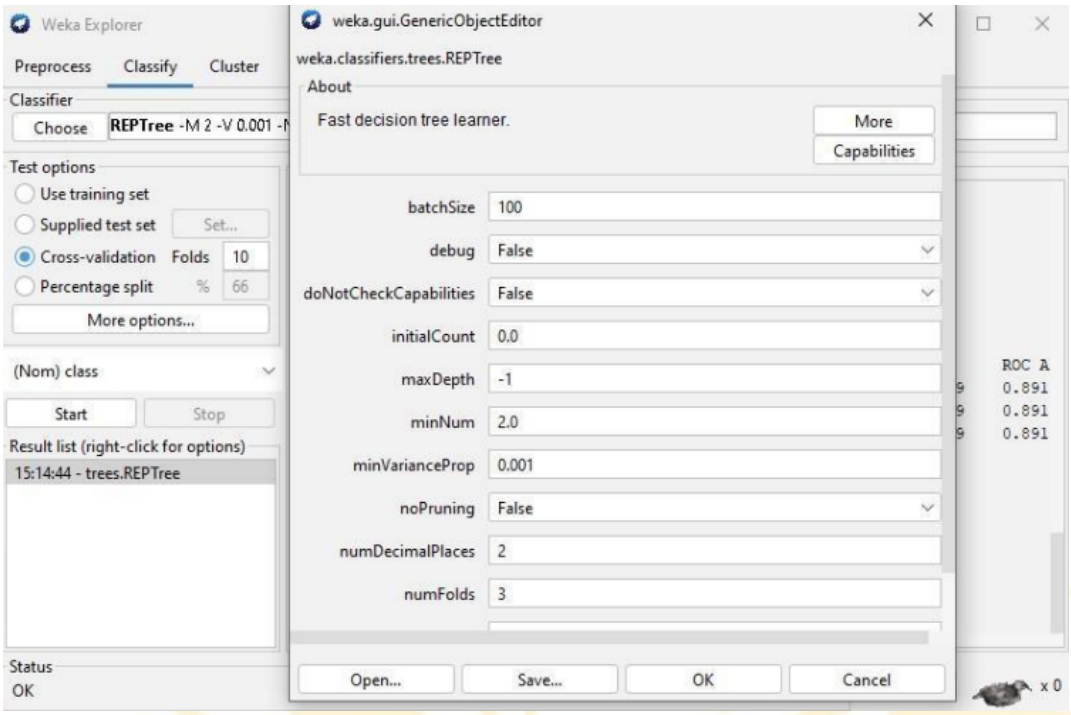
Step 4: Click on the “Classify” tab on the top and then Click the “Choose” button. From the drop-down list, select “trees” which will open all the tree algorithms select the “RepTree” decision tree.



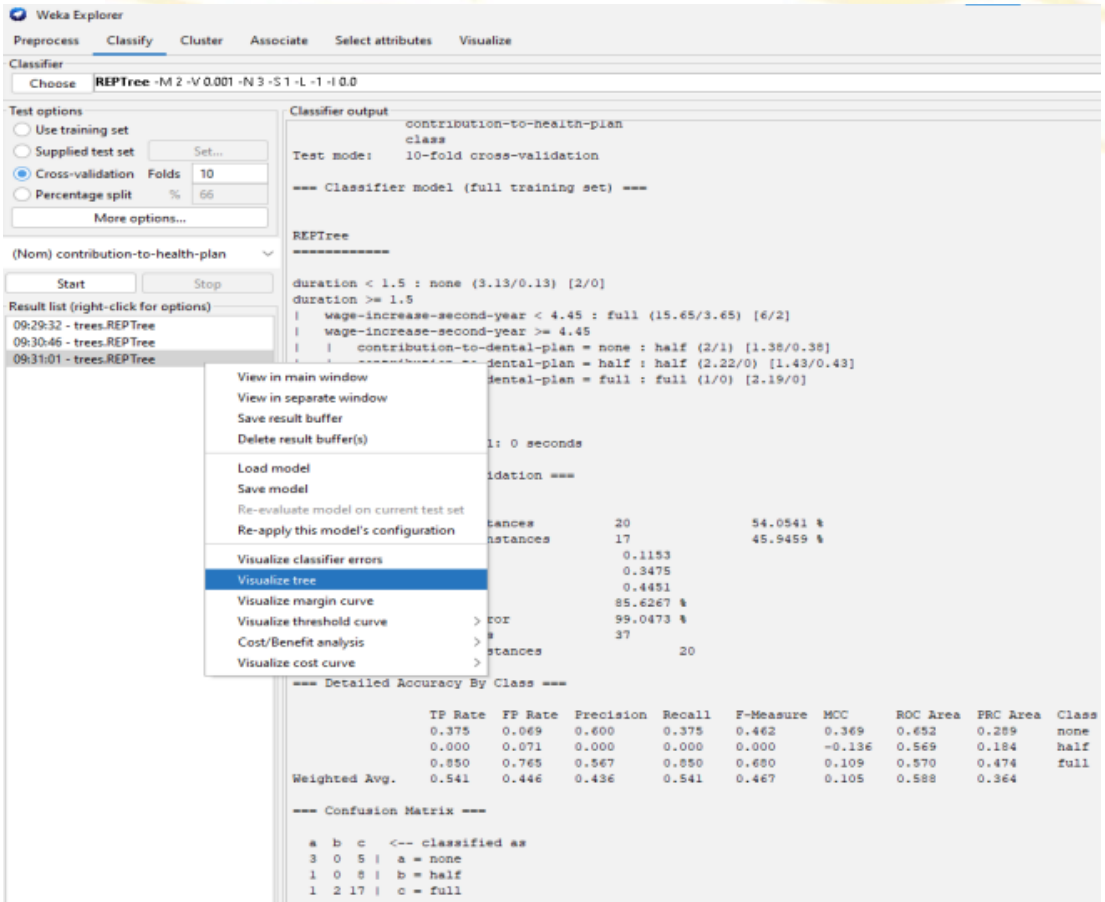
Step 5: Press the “Start” button to start the classification



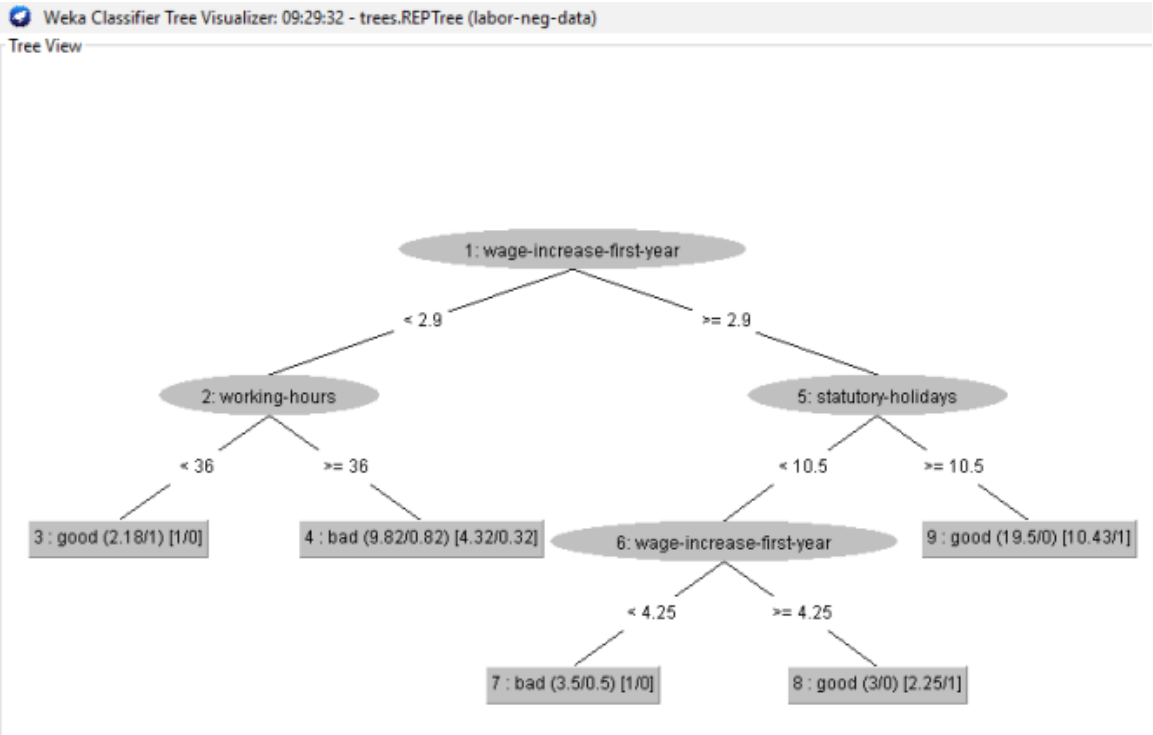
Step 6: We can also access these parameters by clicking on your decision tree algorithm on top



Step 7: Go to the “Result list” section and right-click on your trained algorithm and choose the “Visualize tree” option



Step 8: The tree will look like:



Conclusion:

Successfully implemented K-medoid clustering algorithm using Python.