CLASSIFICATION: - It is necessary to provide a clear classification of data mining systems which may help users to distinguish between such systems and to identify them.

Data mining classification can be done in different ways:

1) Data mining can be classified according to the kinds of databases mined
2) Data mining can be Classified according to the kinds of knowledge mined which is done based on the mining functionalities like characterization, discrimination etc…
3) We can also classify the data mining systems according to the kinds of techniques utilized, applications adapted.

Decision Tree Algorithm: -Classification of data can be done by using decision tree algorithms like J48.

J48 Algorithm:

It is a version of an earlier algorithm named as C4.5 **decision tree algorithm** which is a classic way to represent information from machine learning algorithm.

J48 algorithm gives several options related to tree pruning. The basic idea of the algorithm is to recursively classify the data until each leaf is pure.

J48 uses mainly two pruning methods namely

1. Sub-tree replacement: Where the nodes in a decision tree are replaced with a leaf. This process starts from leaf towards the root.
2. Sub-tree raising: Where nodes may be moved upwards towards the root of the tree, replacing the other nodes along the way.

Now Open the WEKA GUI Chooser from start menu→all programs and click on the EXPLORER button.

Now click on the **Open File** button and choose the file named as "classify.csv" where the content of classify.csv is as shown in the figure 1.

| customer ID	age	income	student	credit rating	class By computer
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle	medium	no	excellent	yes
13	middle	high	yes	fair	yes
14	senior	Medium	no	excellent	no

FIGURE 1: INPUT FILE (CLASSIFY.CSV)
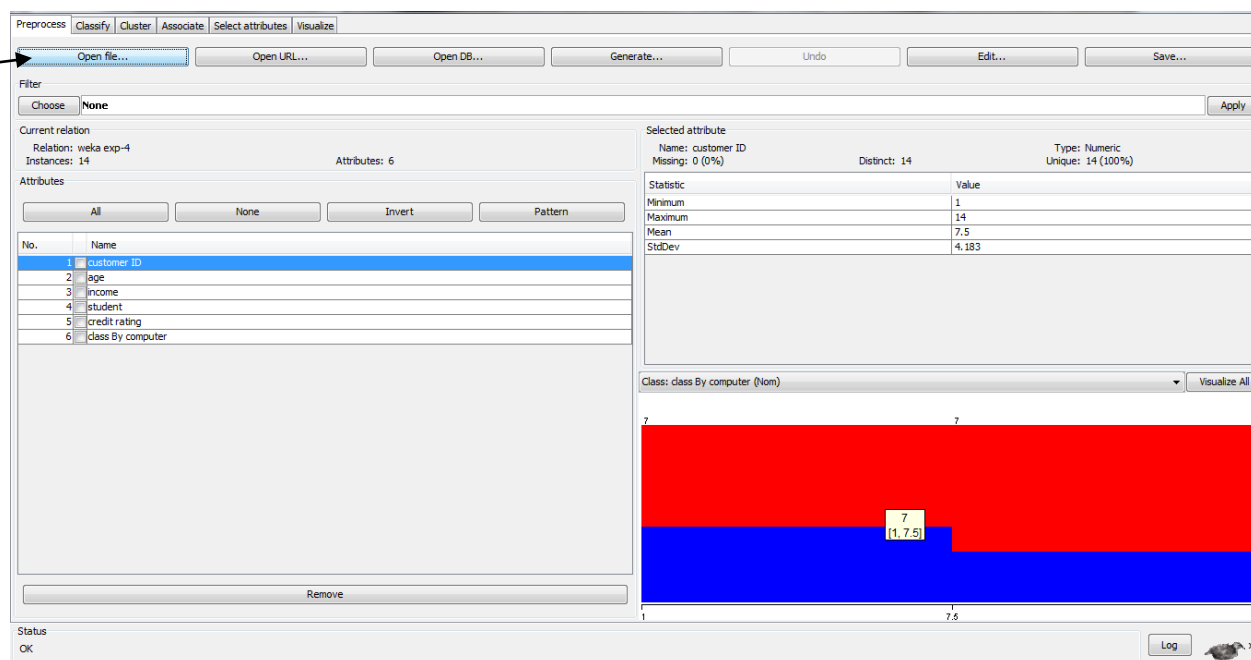
FIGURE 2: LOAD THE FILE CLASSIFY.CSV

After loading the input file named classify.csv as shown in figure 2, choose the classify tab in the WEKA explorer window.

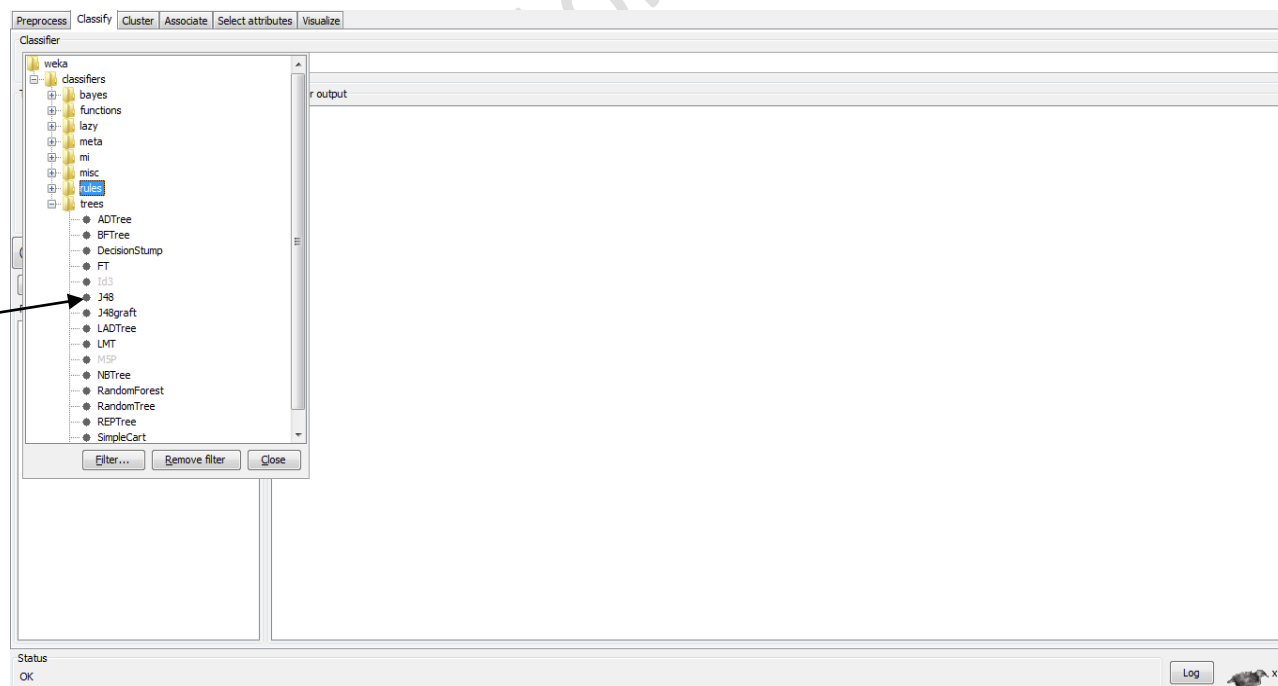Under the classify tab click on choose button and select the J48 under tree as shown in the following figure 3.

Figure 3: SELECTING J48 ALGORITHM (Decision Tree Algorithm)

Now select the "**use training set** "under the **Test Options** located at the left of the WEKA explorer window and click on **start** button.

The output is represented in the **Classifier Output** window in weka explorer window, which is as shown in figure 4 below:
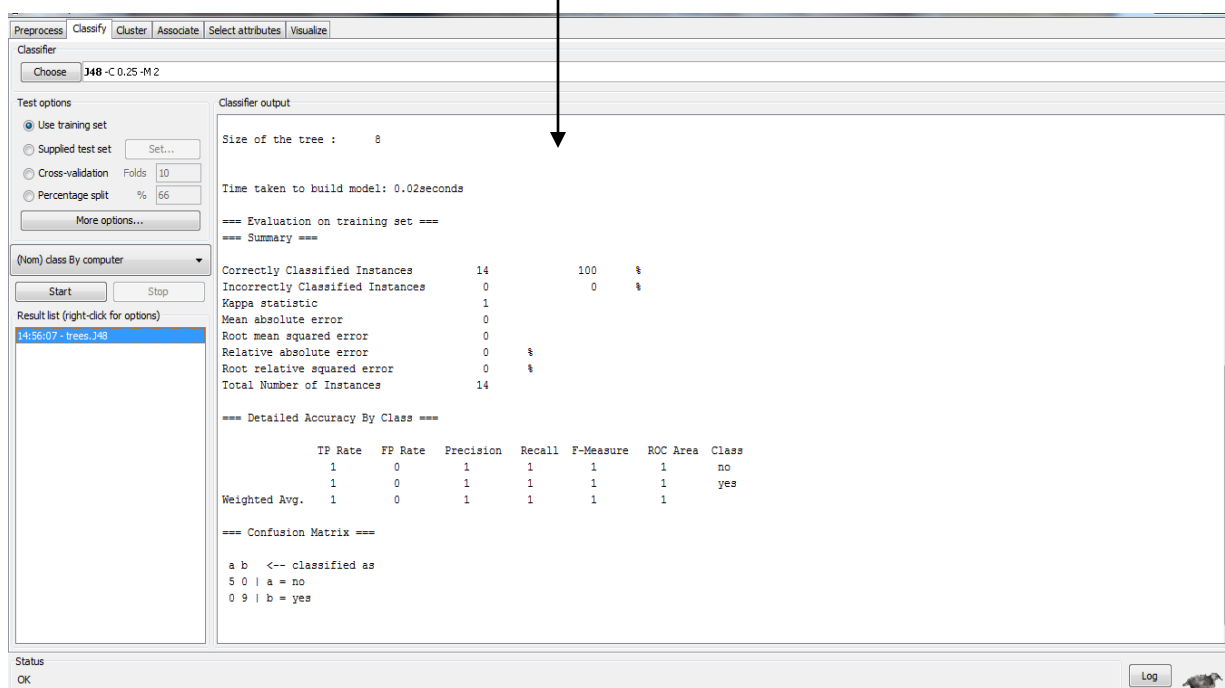

FIGURE 4: SHOWING OUTPUT IN CLASSIFIER OUTPUT WINDOW IN WEKA EXPLORER WINDOW

We can also view the output in a separate window by right clicking on the option in **Result list** and clicking on "**view in separate window**" as shown in figure 5**.**
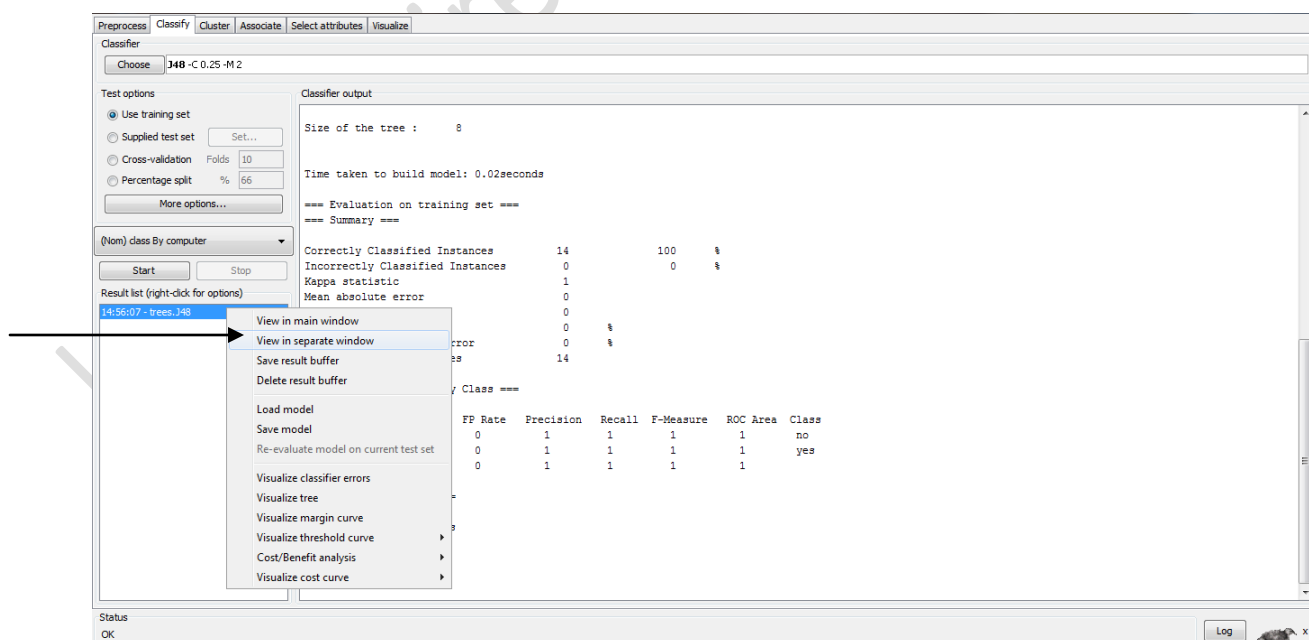

FIGURE 5: SELECTING THE OPTION "VIEW SEPARATE WINDOW"

The output is viewed in a separate window as shown in figure 6 below:

```
14:38:26 - trees.J48

=== Run information ===

Scheme:weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: weka exp-4
Instances:14
Attributes:6
            customer ID
            age
            income
            student
            credit rating
            class By computer
Test mode:evaluate on training data

=== Classifier model (full training set) ===

J48 pruned tree
------------------

age = youth
|   student = no: no (3.0)
|   student = yes: yes (2.0)
age = middle: yes (4.0)
age = senior
|   credit rating = fair: yes (3.0)
|   credit rating = excellent: no (2.0)

Number of Leaves  :     5

Size of the tree :      8


Time taken to build model: 0.05seconds

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances        14               100      %
Incorrectly Classified Instances       0                 0      %
Kappa statistic                        1
Mean absolute error                    0
Root mean squared error                0
Relative absolute error                0       %
Root relative squared error            0       %
Total Number of Instances             14

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
                 1        0         1         1        1          1       no
                 1        0         1         1        1          1       yes
Weighted Avg.    1        0         1         1        1          1

=== Confusion Matrix ===

 a b   <-- classified as
 5 0 | a = no
 0 9 | b = yes
```

FIGURE 6: OUTPUT SCREEN IN A SEPARATE WINDOW

Under the **Result List** right click the item to get the options as shown in the figure 7 and select the option "**visualize tree**" option.
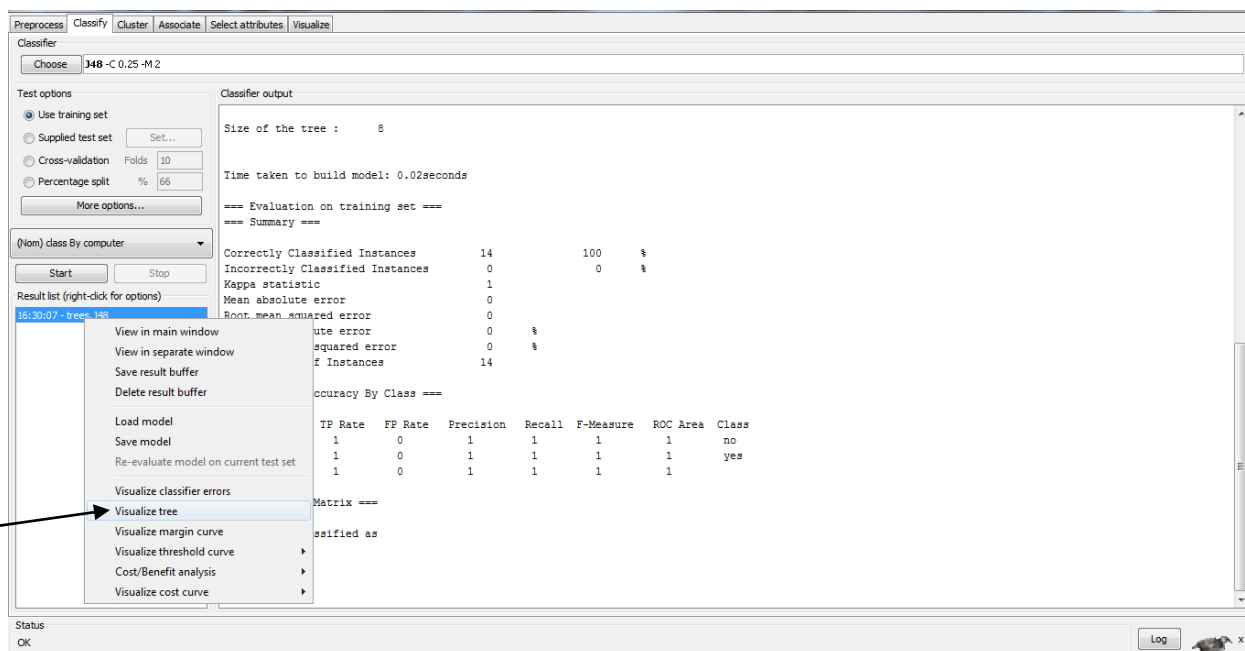

FIGURE 7: OUTPUT SCREEN SHOWS HOW TO SELECT VISUALIZE AS TREE OPTION

After selecting the" **visualize tree**" option the output is represented as a tree in a separate window as shown in the figure 8 given below.
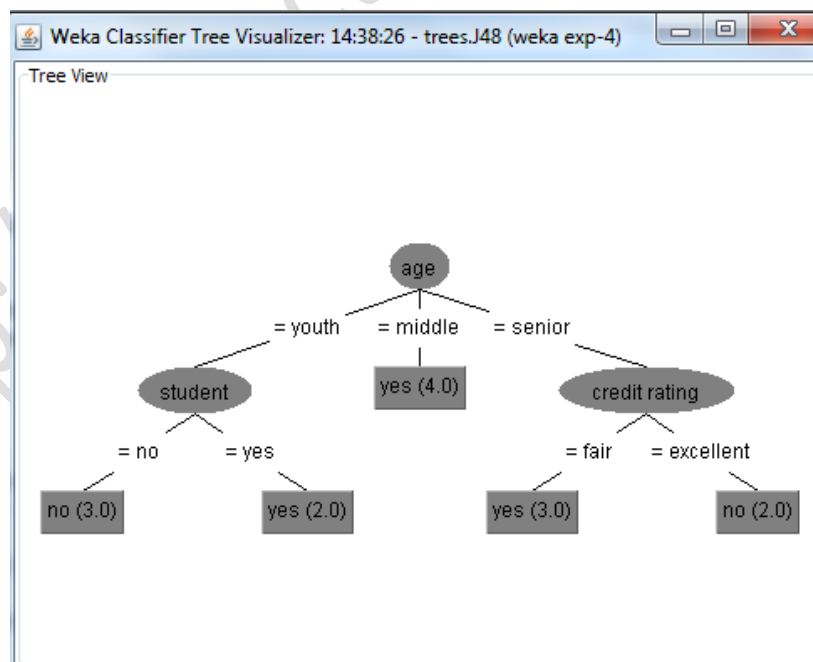

FIGURE 8: TREE OUTPUT OF CLASSIFICATION