EX.NO:27
Date:

PREDICTION OF CATEGORICAL DATA USING DECISION TREE ALGORTIHM THROUGH WEKA

AIM:

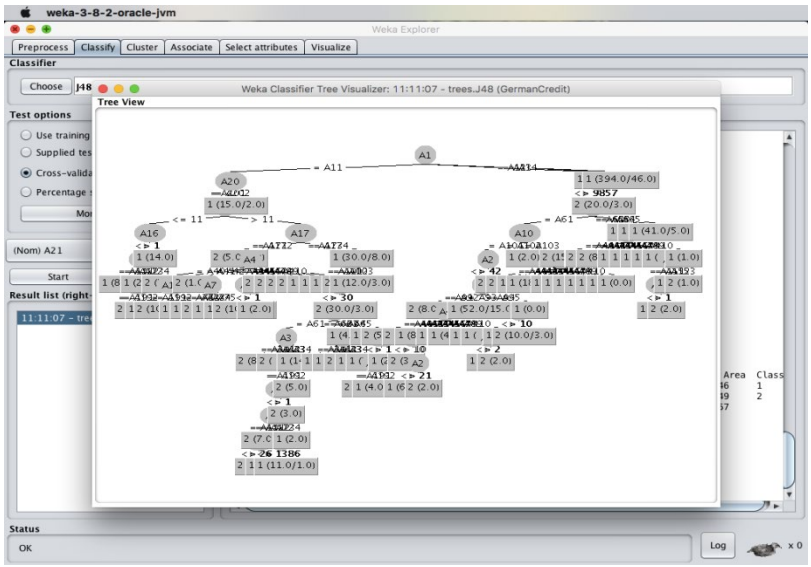
To create prediction of categorical data using decision tree algorithm through weka tool.

PROCEDURE:

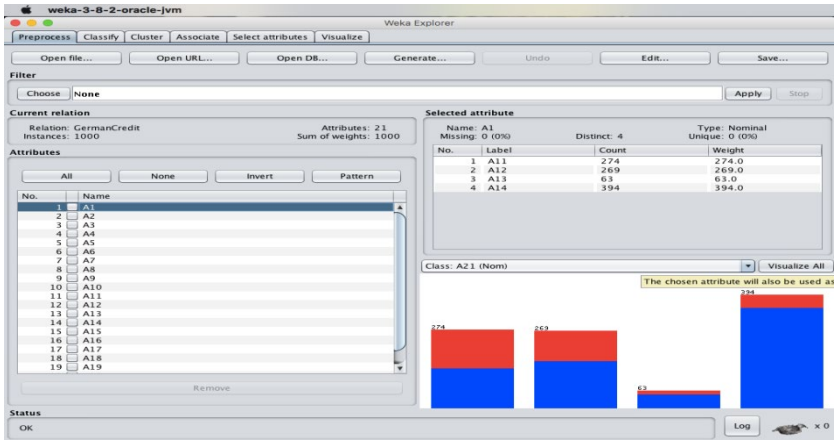
1.Download WEKA And Install
2.Start WEKA
3.Open The Data/iris.arff Dataset
4.Select And Run An Algorithm
5.Review The Results

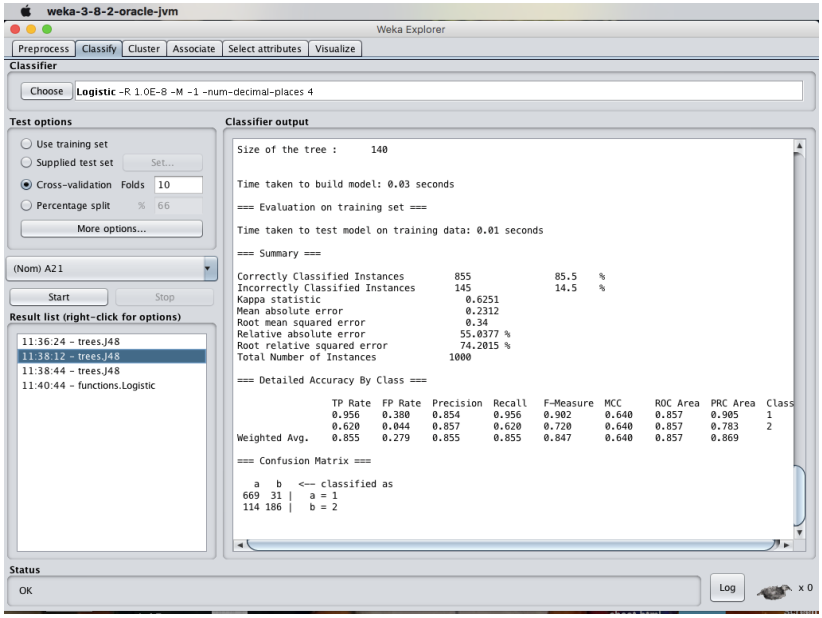
❖ Decision Tree :

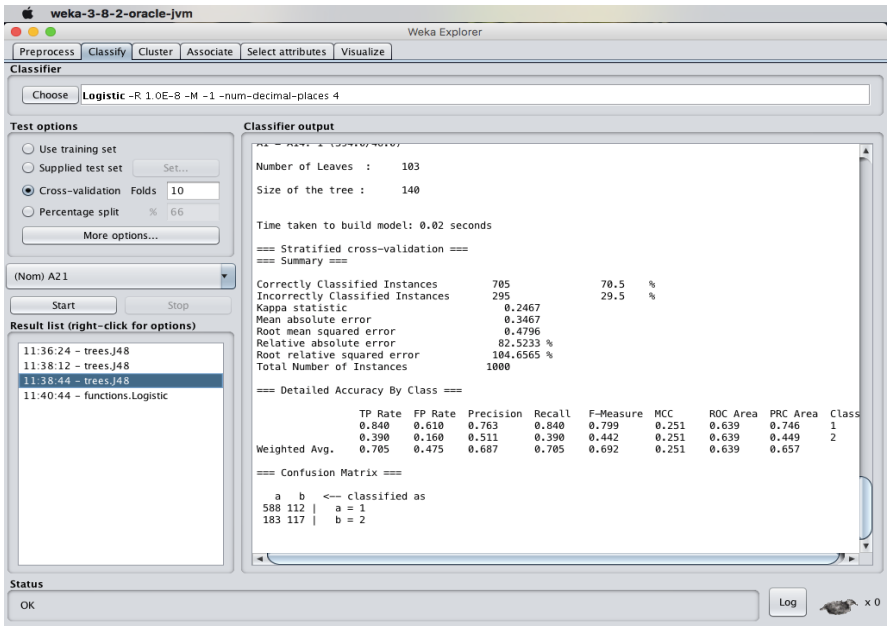
Visualize the decision tree for the given dataset.



or.

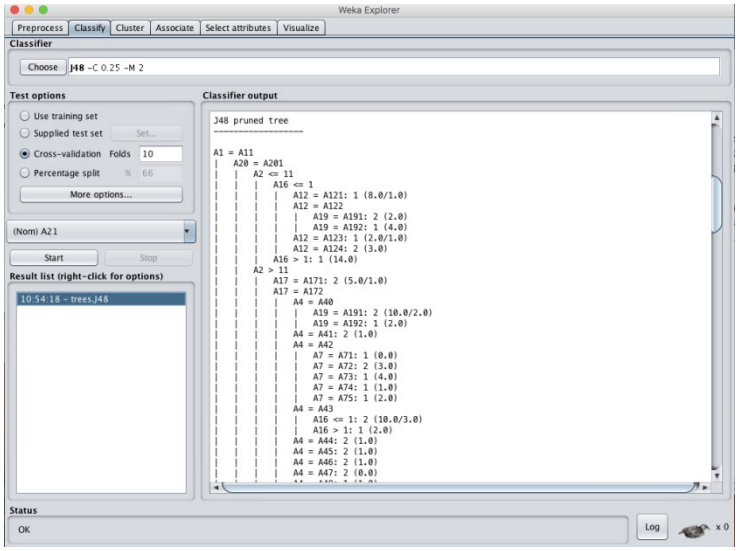

weka-3-8-2-oracle-jvm
Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
None
Apply
Stop
Current relation
Relation: GermanCredit
Instances: 1000
Attributes: 21
Sum of weights: 1000
Selected attribute
Name: A1
Missing: 0 (0%)
Distinct: 4
Type: Nominal
Unique: 0 (0%)
No. Label Count Weight
1 A11 274 274.0
2 A12 269 269.0
3 A13 63 63.0
4 A14 394 394.0
Attributes
All
None
Invert
Pattern
No. Name
Remove
Class: A21 (Nom)
Visualize All
The chosen attribute will also be used as
Status
OK
Log
x 0

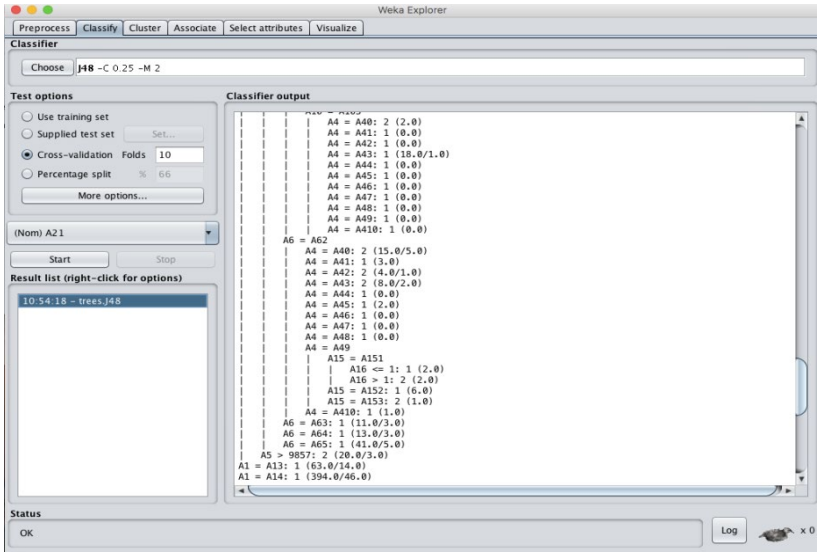

weka-3-8-2-oracle-jvm
Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
Logistic -R 1.0E-8 -M -1 -num-decimal-places 4
Test options
Use training set
Supplied test set
Set...
Cross-validation Folds 10
Percentage split % 66
More options...
(Nom) A21
Start
Stop
Result list (right-click for options)
11:36:24 - trees.J48
11:38:12 - trees.J48
11:38:44 - trees.J48
11:40:44 - functions.Logistic
Classifier output
Size of the tree : 140
Time taken to build model: 0.03 seconds
=== Evaluation on training set ===
Time taken to test model on training data: 0.01 seconds
=== Summary ===
Correctly Classified Instances 855 85.5 %
Incorrectly Classified Instances 145 14.5 %
Kappa statistic 0.6251
Mean absolute error 0.2312
Root mean squared error 0.34
Relative absolute error 55.0377 %
Root relative squared error 74.2015 %
Total Number of Instances 1000
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class
0.956 0.380 0.854 0.956 0.902 0.640 0.857 0.905 1
0.620 0.044 0.857 0.620 0.720 0.640 0.857 0.783 2
Weighted Avg. 0.855 0.279 0.855 0.855 0.847 0.640 0.857 0.869
=== Confusion Matrix ===
a b <-- classified as
669 31 | a = 1
114 186 | b = 2
Status
OK
Log
x 0

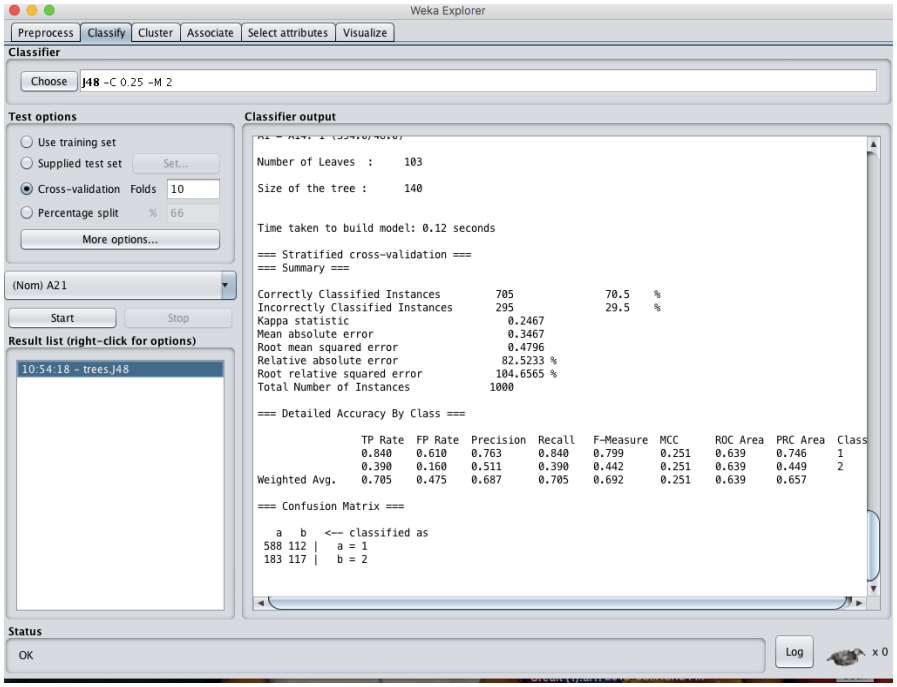

weka-3-8-2-oracle-jvm
Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
Logistic -R 1.0E-8 -M -1 -num-decimal-places 4
Test options
Use training set
Supplied test set
Set...
Cross-validation Folds 10
Percentage split % 66
More options...
(Nom) A21
Start
Stop
Result list (right-click for options)
11:36:24 - trees.J48
11:38:12 - trees.J48
11:38:44 - trees.J48
11:40:44 - functions.Logistic
Classifier output
Number of Leaves : 103
Size of the tree : 140
Time taken to build model: 0.02 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances 705 70.5 %
Incorrectly Classified Instances 295 29.5 %
Kappa statistic 0.2467
Mean absolute error 0.3467
Root mean squared error 0.4796
Relative absolute error 82.5233 %
Root relative squared error 104.6565 %
Total Number of Instances 1000
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class
0.840 0.610 0.763 0.840 0.799 0.251 0.639 0.746 1
0.390 0.160 0.511 0.390 0.442 0.251 0.639 0.449 2
Weighted Avg. 0.705 0.475 0.687 0.705 0.692 0.251 0.639 0.657
=== Confusion Matrix ===
a b <-- classified as
588 112 | a = 1
183 117 | b = 2
Status
OK
Log
x 0

➢ CROSS VALIDATION ANALYSIS :

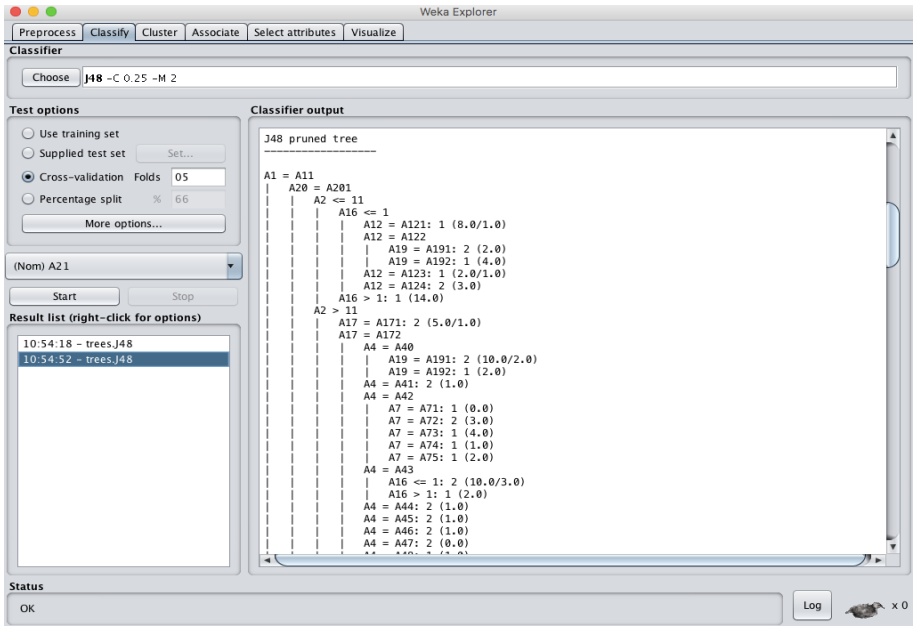
When cross validation folds are 10 :

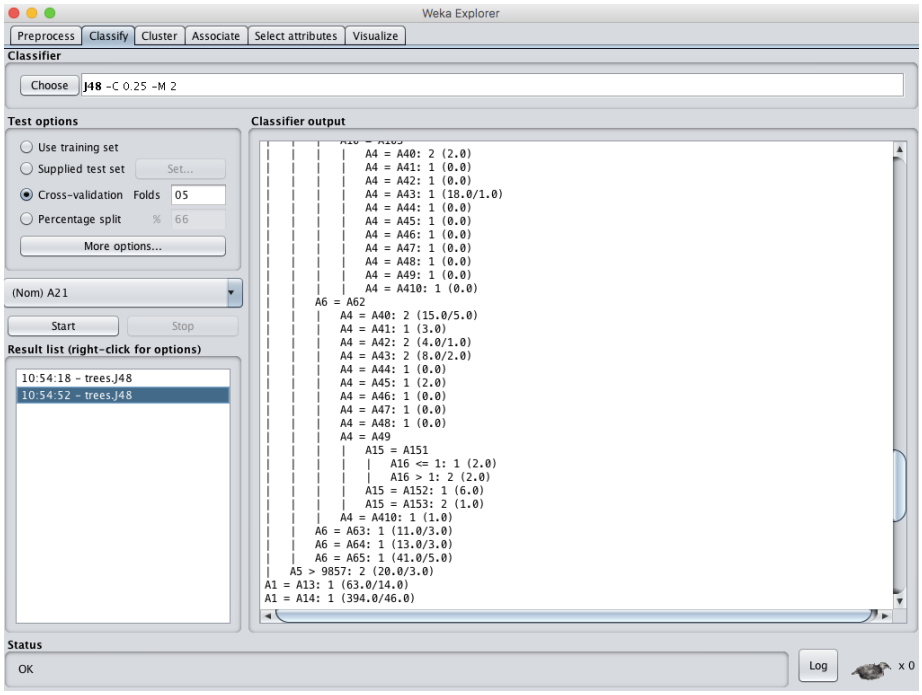


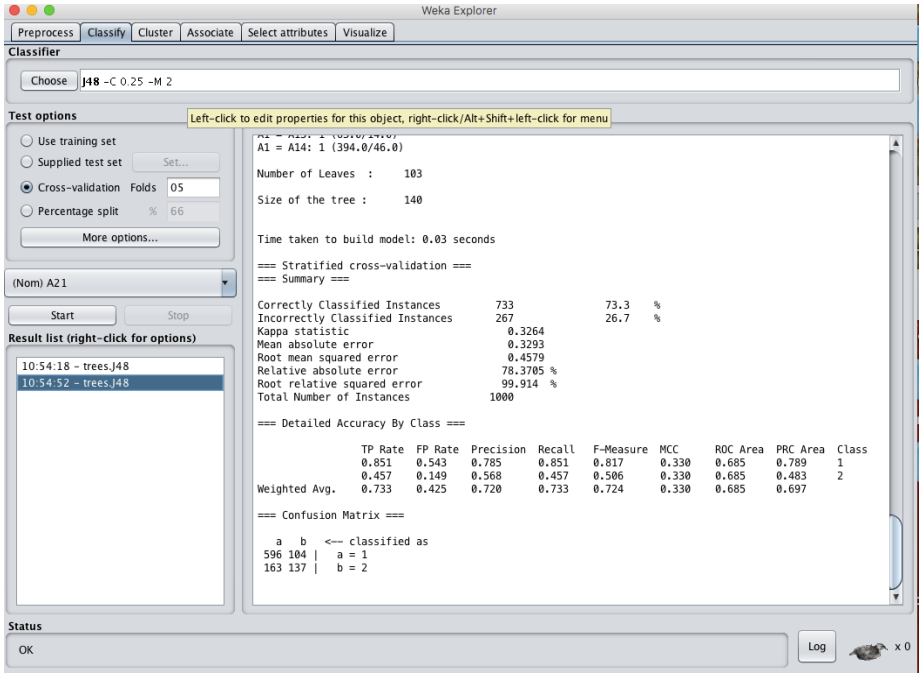




When cross validation folds are : 05 :-







RESULT :

Thus, the observations and evaluations done on the german_credit dataset are analyzed. The decision tree has been successfully visualized. Various evaluations and comparisons done through the cross validation folds change. Which lead to the change of values in confusion matrix.