S.NO	EXPERIMENT NAME
21.	**DATA PREPROCESSING AND ANALYSIS FOR DATASET USING WEKA**
22.	**DATA SEGMENTATION BY K- MEANS CLUSTER USING WEKA AND R-TOOL**
23.	**DATASEGMENTATION BY EXPECTATION MAXIMISATION ALGORITHM THROUGH WEKA**
24.	**DATA SEGMENTATION BY COBWEB – HIERARCHIAL CLUSTERING ALGORITHM USING WEKA TOOL**
25.	**FREQUENT PATTERN MINING USING ASSOCIATION RULE THROUGH WEKA AND R TOOLS**
26.	**FREQUENT PATTERN MINING USING FP GROWTH THROUGH WEKA TOOL**
27.	**PREDICTION OF CATEGORICAL DATA USING DECISION TREE ALGORTIHM THROUGH WEKA**
28.	**PREDICTION OF CATEGORICAL DATA USING SMO ALGORTIHM THROUGH WEKA**
29.	**EVALUATING ACCURACY OF THE CLASSIFIERS**
30.	**DESCRIPTION NUMERICAL PREDICTION ANALYSIS USING LINEAR REGRESSION THROUGH WEKA**

EX.NO : 21
Date :

DATA PREPROCESSING AND ANALYSIS FOR DATASET USING WEKA

AIM:

TO Create data preprocessing and analysis for dataset using weka.

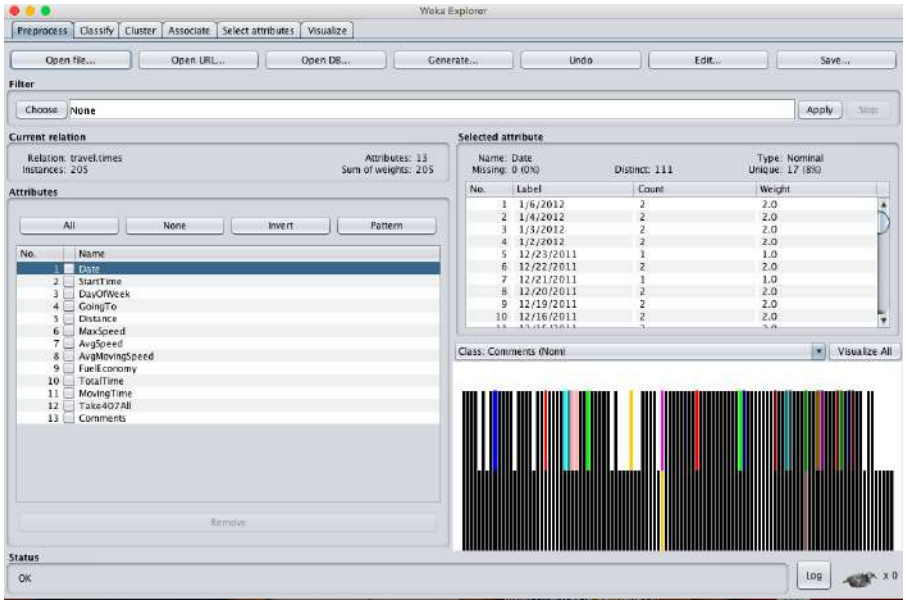
DESCRIPTION:

Consider a dataset of traveltimes.csv file where it contains the columns of (or) attributes as Date, StartTime, DayOfWeek, GoingTo, Distance, MaxSpeed, AvgSpeed, AvgMovingSpeed, FuelEconomy, TotalTime, MovingTime, Take407All comments.

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

PREPROCESS:



OBSERVATION :

A. ATTRIBUTE TYPE :

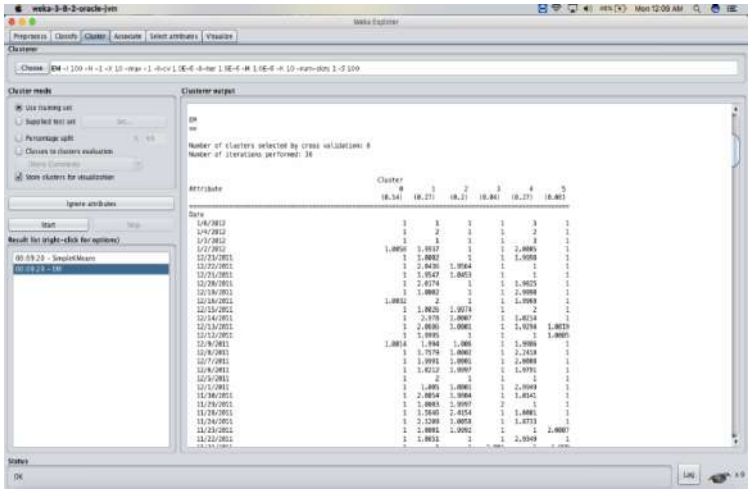
S.NO	ATTRIBUTE	TYPE
1.	Date	Nominal
2.	Start Time	Nominal
3.	Day Of Week	Nominal
4.	Going To	Nominal
5.	Distance	Numeric
6.	Max Speed	Numeric
7.	Avg Speed	Numeric
8.	Avg Moving Speed	Numeric
9.	Fuel Economy	Nominal

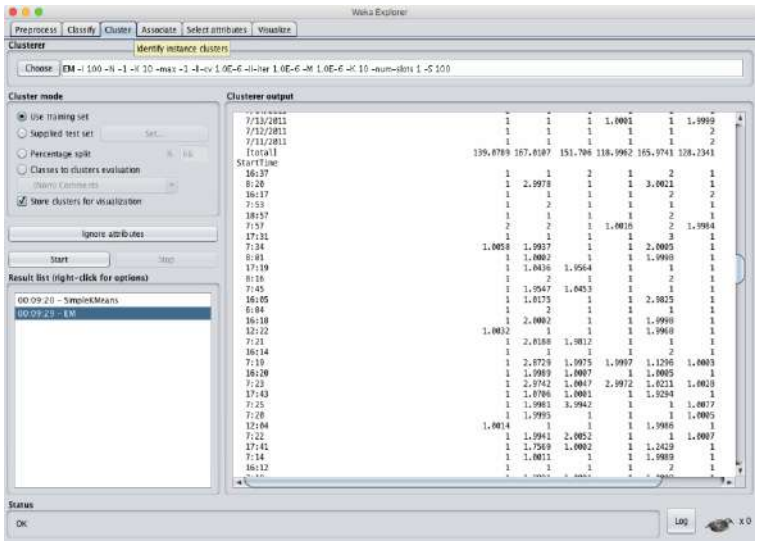
10.	Total Time	Numeric
11.	Moving Time	Numeric
12.	Comments	Nominal
13.	Take 407 All	Nominal

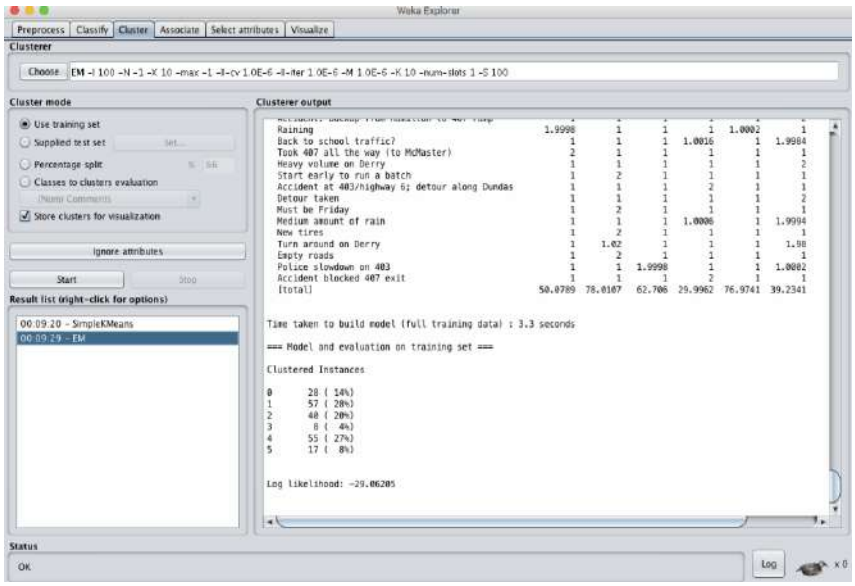
B.PERCENTAGE OF MISSING VALUES :

S.NO	ATTRIBUTE	Percentage Of Missing Values
1.	Date	0 %
2.	Start Time	0 %
3.	Day Of Week	0 %
4.	Going To	0 %
5.	Distance	0 %
6.	Max Speed	0 %
7.	Avg Speed	0 %
8.	Avg Moving Speed	0 %
9.	Fuel Economy	8 %
10.	Total Time	0 %
11.	Moving Time	0 %
12.	Comments	88 %
13.	Take 407 All	0 %

B. MIN, MAX, MEAN, STANDARD DEVIATION :







RESULT :

Thus, the dataprocessing and analysis for a dataset using weka tool has been successfully completed.

EX.NO.22:
Date:

DATA SEGMENTATION BY K-MEANS CLUSTER USING WEKA AND R-TOOL

AIM:

To create DataSegmentation by k-means cluster using weka and R-tool.

DESCRIPTION:

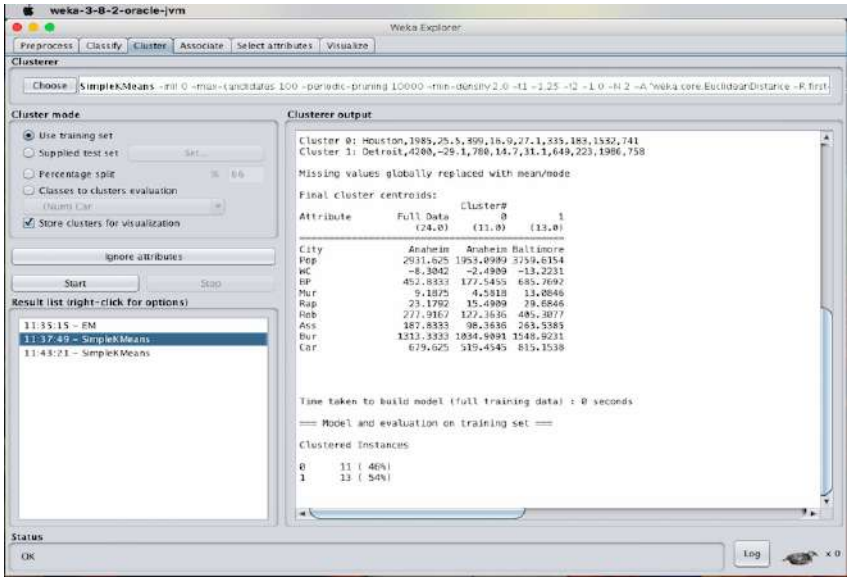
Consider a dataset of citycrimes.csv file of which it contains the attributes are City, Pop, WC, BP, Mur, Rap, Rob, Ass, Bus and car for the performance of the dataset by applying the K-means algorithm in weka and as well using R- tool.

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.

USING WEKA TOOL :
STEPS INVOLVED :

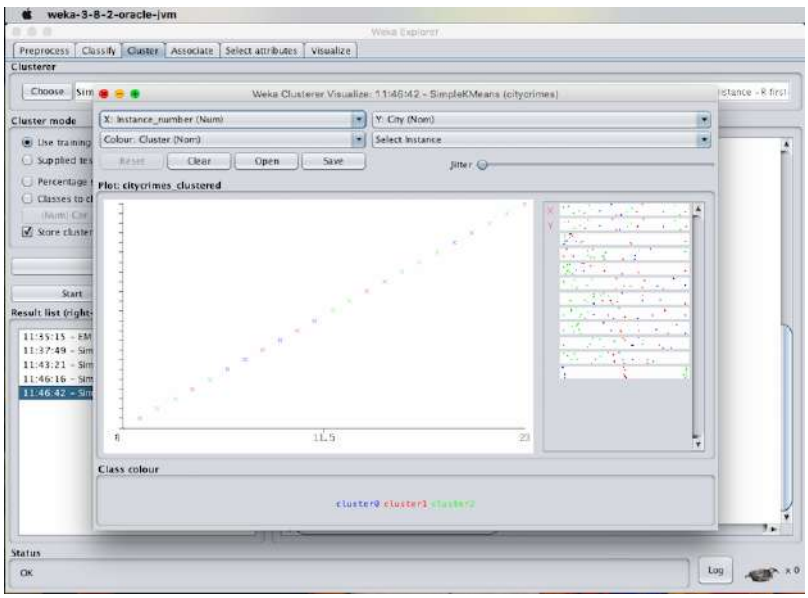
- Choose a set of attributes for clustering and for giving a motivation.
- Choose the dataset and import the dataset into Weka tool.
- Cluster the dataset and choose simple K-means algorithm and give the motivation.



A. Experiment with atleast 2 different number of clusters but with same seed values:

STEPS INVOLVED :

- Compare the two different clusters but with the same seed values.
- Change the number of clusters value and need not to change the seed value.
- Apply the K-means algorithm and start executing the algorithm.



RESULT :

Thus, the K-means clustering analyzing using the weka tool has been successfully completed. In case of weka tool, the change in seed values lead to the decrease in the number of iterations.

EX.NO:23
Date:

DATA SEGMENTATION BY EXPECTATION MAXIMISATION ALGORITHM THROUGH WEKA

AIM:

To create data segmentation by Expectation Maximisation algorithm through weka.

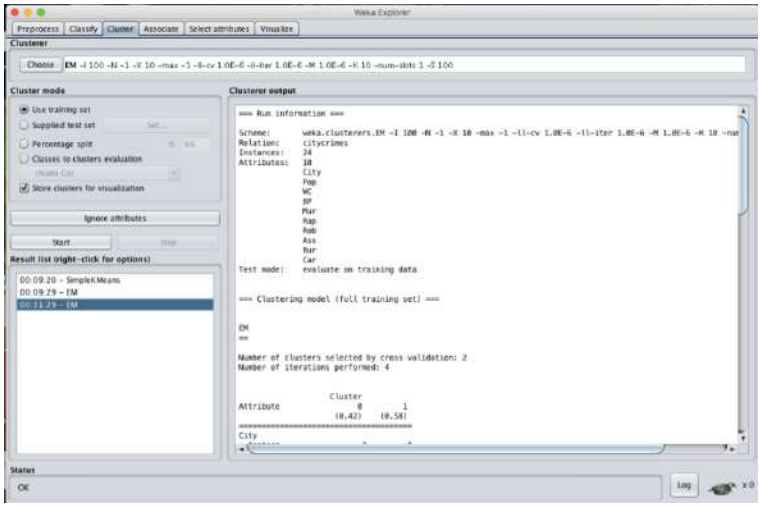
DESCRIPTION:

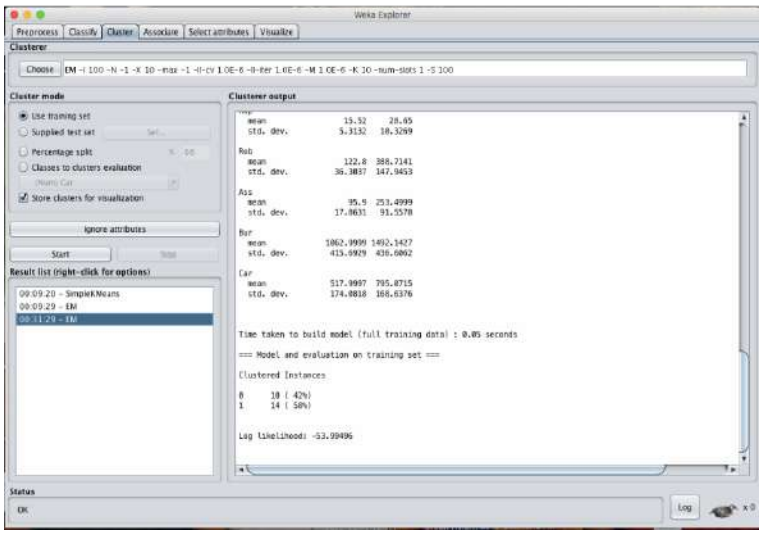
Consider a dataset of citycrimes.csv file of which it contains the attributes are City, Pop, WC, BP, Mur, Rap, Rob, Ass, Bus and car for the performance of the dataset by applying the K-means algorithm in weka and as well using R- tool.

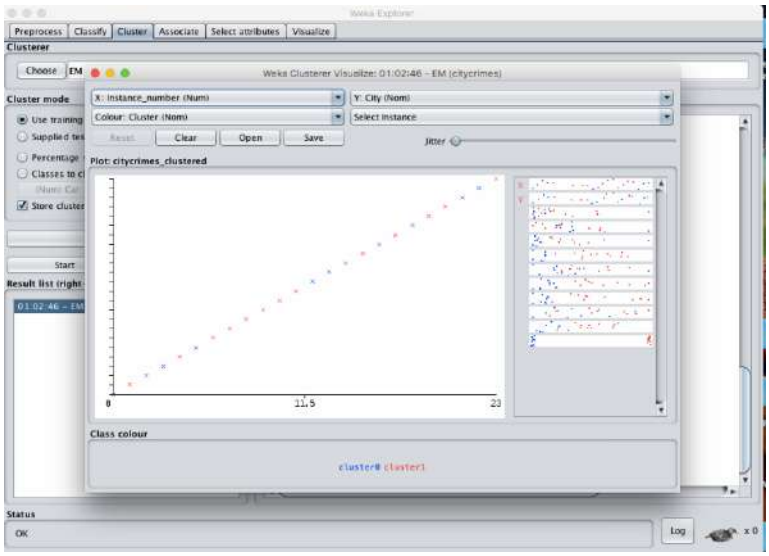
When the clustering is been made through the expectation maximization algorithm by setting minimum standard deviation values then the results will be of the following :

PROCEDURE STEPS:

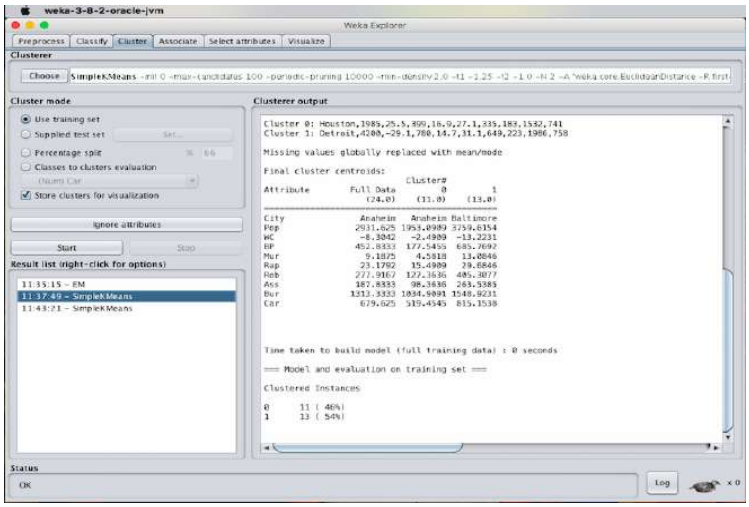
- Initially, load the dataset into the weka tool and check for all the attributes present in the dataset.
- Then move to cluster panel and apply the EM algorithm technique for the datasheet.
- Finally, Observe the results that are obtained.


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
Choose
EM -I 100 -N -1 -X 10 -max -1 -ll-cv 1.0E-6 -ll-iter 1.0E-6 -M 1.0E-6 -K 10 -num-slots 1 -S 100
Cluster mode
Use training set
Supplied test set
Percentage split
Classes to clusters evaluation
Store clusters for visualization
Ignore attributes
Start
Result list (right-click for options)
00:09:20 - SimpleKMeans
00:09:29 - EM
00:11:29 - EM
Clusterer output
=== Run information ===
Scheme: weka.clusterers.EM -I 100 -N -1 -X 10 -max -1 -ll-cv 1.0E-6 -ll-iter 1.0E-6 -M 1.0E-6 -K 10 -num
Relation: citycrimes
Instances: 24
Attributes: 10
City
Pop
WC
BP
Mur
Rap
Rob
Ass
Bur
Car
Test mode: evaluate on training data
=== Clustering model (full training set) ===
EM
==
Number of clusters selected by cross validation: 2
Number of iterations performed: 4
Cluster
Attribute 0 1
(0.42) (0.58)
City
Status
OK
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
Choose
EM -I 100 -N -1 -X 10 -max -1 -ll-cv 1.0E-6 -ll-iter 1.0E-6 -M 1.0E-6 -K 10 -num-slots 1 -S 100
Cluster mode
Use training set
Supplied test set
Percentage split
Classes to clusters evaluation
Store clusters for visualization
Ignore attributes
Start
Result list (right-click for options)
00:09:20 - SimpleKMeans
00:09:29 - EM
00:11:29 - EM
Clusterer output
mean 15.52 28.05
std. dev. 5.3132 10.3289
Rob
mean 122.8 388.7141
std. dev. 36.3837 147.9453
Ass
mean 95.9 253.4999
std. dev. 17.0631 91.5578
Bur
mean 1062.9999 1492.1427
std. dev. 415.9929 436.6092
Car
mean 517.9997 795.8715
std. dev. 174.0818 188.6376
Time taken to build model (full training data) : 0.05 seconds
=== Model and evaluation on training set ===
Clustered Instances
0 10 (42%)
1 14 (58%)
Log likelihood: -53.99496
Status
OK
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
Choose
EM
Weka Clusterer Visualize: 01:02:46 - EM (citycrimes)
X: instance_number (Num)
Y: City (Nom)
Colour: Cluster (Nom)
Select Instance
Reset
Clear
Open
Save
Jitter
Plot: citycrimes_clustered
0
11.5
23
Class colour
cluster0 (cluster)
Cluster mode
Result list (right
01:02:46 - EM
Status
OK
Log
x 0

❖ K- MEANS ALGORITHM:



RESULT :

Thus, the data analysis by the expectation maximization algorithm using weka has been analyzed and observed properly.

EX.NO:24
Date:

DATA SEGMENTATION BY COBWEB – HIERARCHIAL CLUSTERING ALGORITHM USING WEKA TOOL

AIM:

To create data segmentation by cobweb-hierarchial clustering algorithm using weka tool.

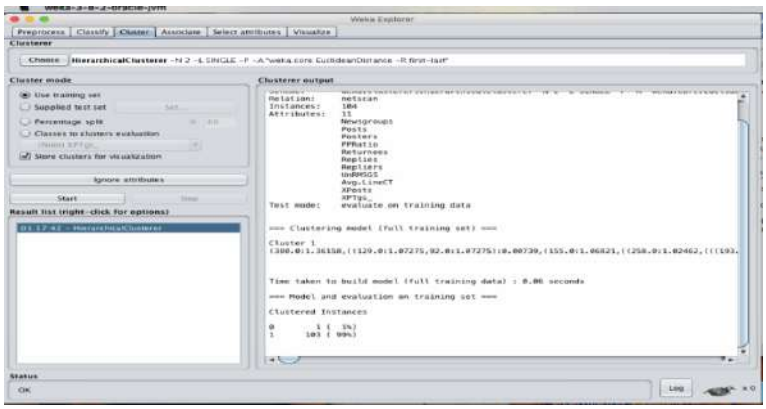
DESCRIPTION:

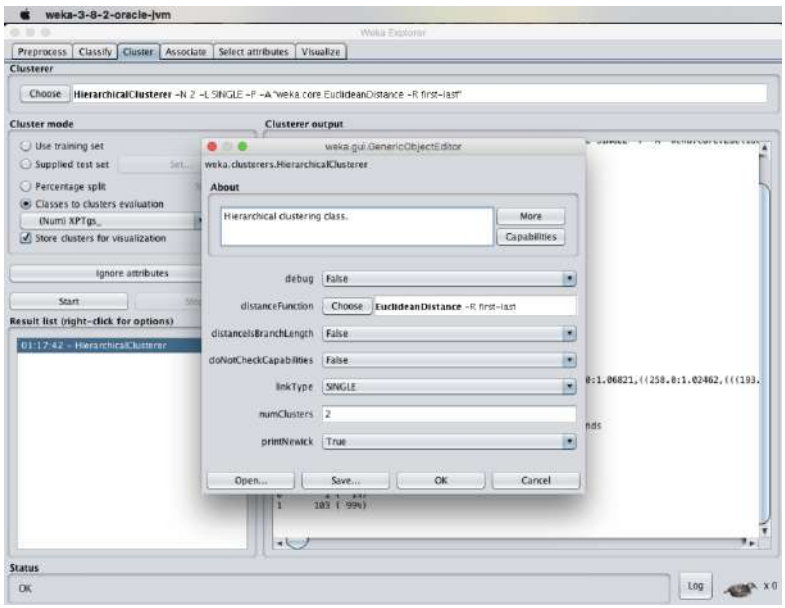
Consider a dataset netscan.csv where it contains the attributes of Newsgroups, posts, posters, PPRatio, Returnees, Replies, Repliers, UnRMsgs, Avg.LineCT, Xports, XPTgs. Each attribute will have different types of the meanings.

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

HIERARCHIIAL CLUSTERING:





RESULT :

Thus, the data analysis of cobweb hierarchial clustering algorithm using weka tools has been analyzed and observed successfully.

EX.NO:25
Date:

FREQUENT PATTERN MINING USING ASSOCIATION RULE THROUGH WEKA AND R -TOOLS

AIM:

To create frequent pattern mining using association rule through weka and R-tools.

DESCRIPTION:

Consider a dataset of 2015.csv file of which it contains the attributes are Reference Number, Grid ref: Easting, Grid Ref: Northing, Number of vehicles, Accident date, Time(24 hr), 1st Road class, Road Surface, Lighting conditions, Weather conditions, casuality class, Sex of casuality, Age of casuality, Type of casuality for the performance of the dataset by applying the Apriori algorithm in weka and as well using R- tool.

PROCEDURE:

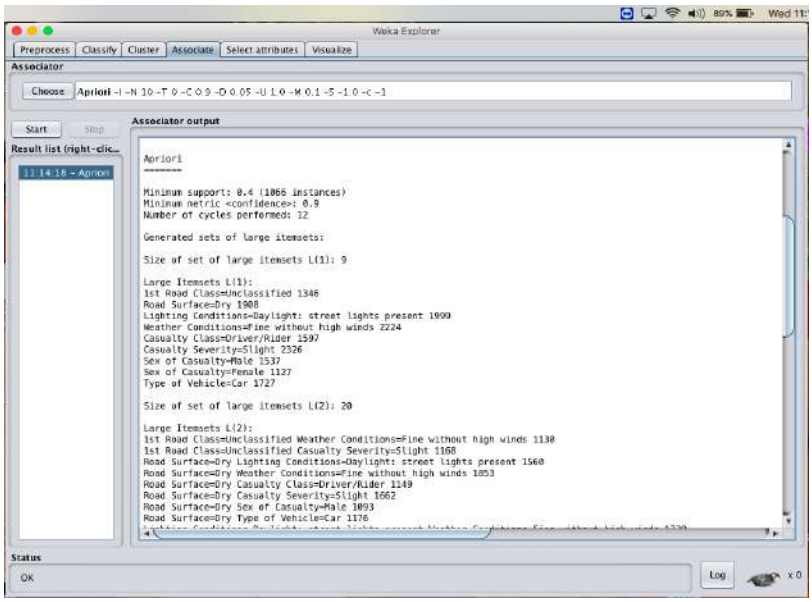
1.Download WEKA And Install
2.Start WEKA
3.Open The Data/iris.arff Dataset
4.Select And Run An Algorithm
5.Review The Results

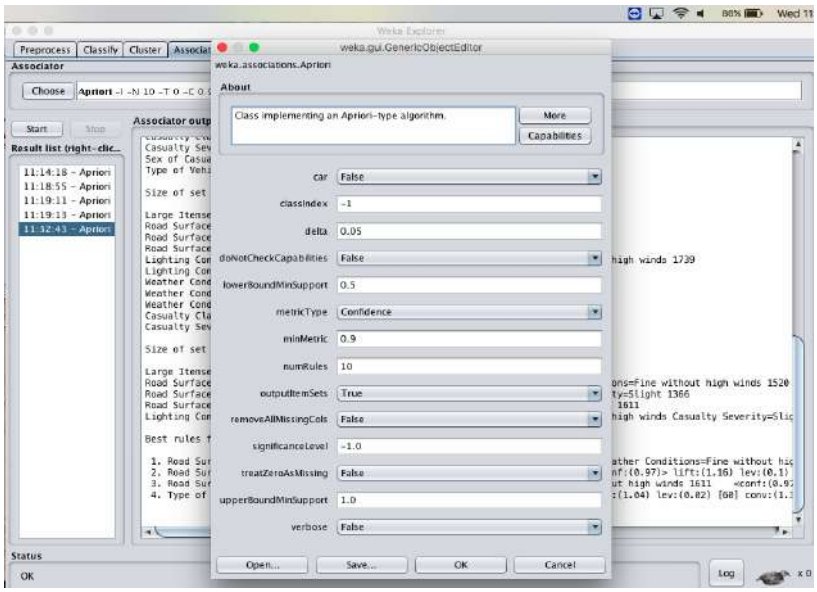
❖ USING WEKA TOOL :

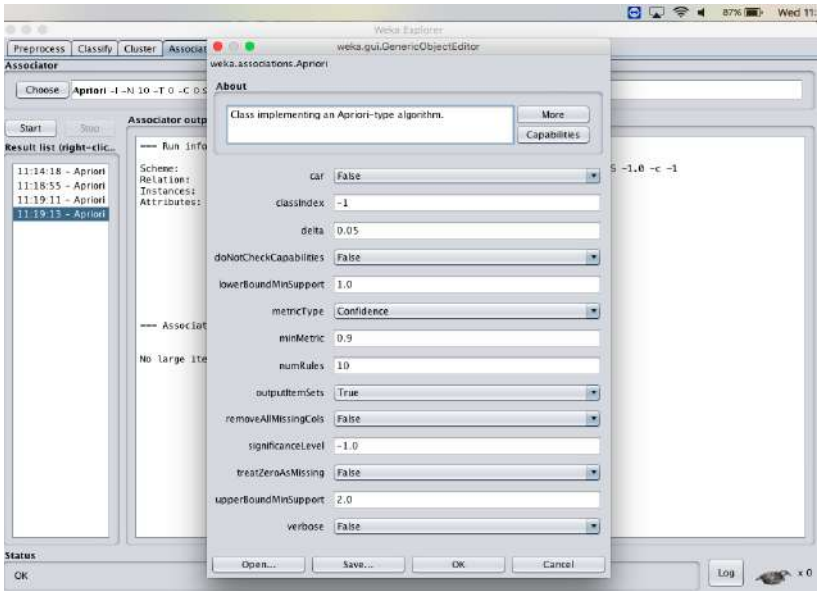
STEPS INVOLVED :

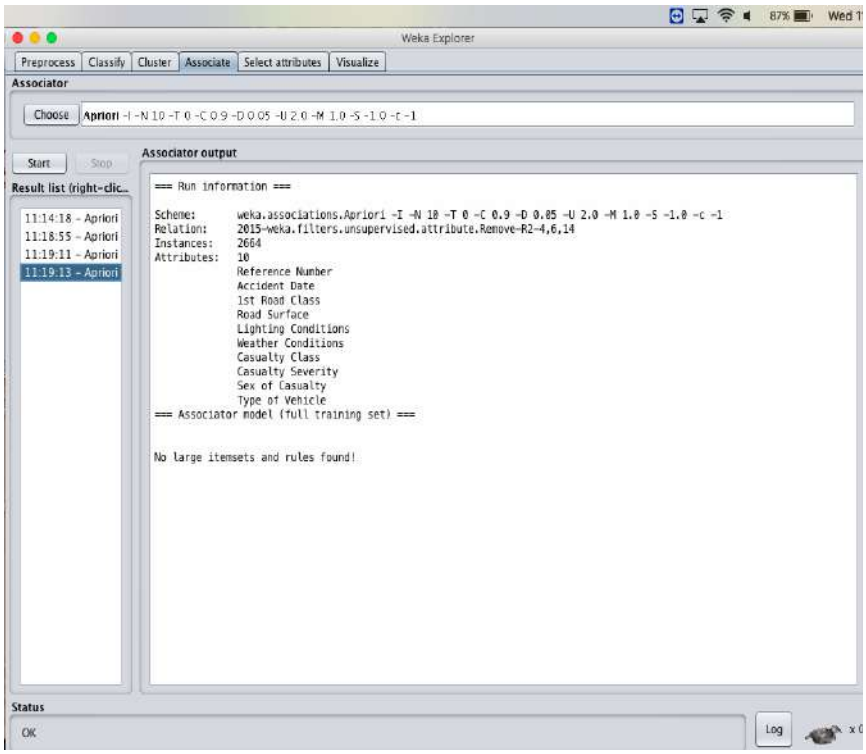
- Choose a set of attributes for clustering and for giving a motivation.
- Choose the dataset and import the dataset into Weka tool.
- Discretize the attributes from numeric to nominal to perform the algorithm.

- Cluster the dataset and choose simple Apriori algorithm.
- Set the Upper bound min_sup and lower bound min_sup values.









RESULT :

Thus, the Apriori algorithm analyzing using both the weka tool and R- tool has been successfully completed. In case of weka tool, the change in upper bound and lower bound values lead to the increase and decrease of number of

itemsets and rules . In case of R-tool, there is an increase in absolute minimum support count value.

EX.NO:26
Date:

FREQUENT PATTERN MINING USING FP GROWTH THROUGH WEKA TOOL

AIM:

To create frequent pattern mining using FP Growth through weka tool.

DESCRIPTION:

Consider a dataset of 2015.csv file of which it contains the attributes are Reference Number, Grid ref: Easting, Grid Ref: Northing, Number of vehicles, Accident date, Time(24 hr), 1st Road class, Road Surface, Lighting conditions, Weather conditions, casuality class, Sex of casuality, Age of casuality, Type of casuality for the performance of the dataset by applying the FP algorithm in 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

- **USING WEKA TOOL :**

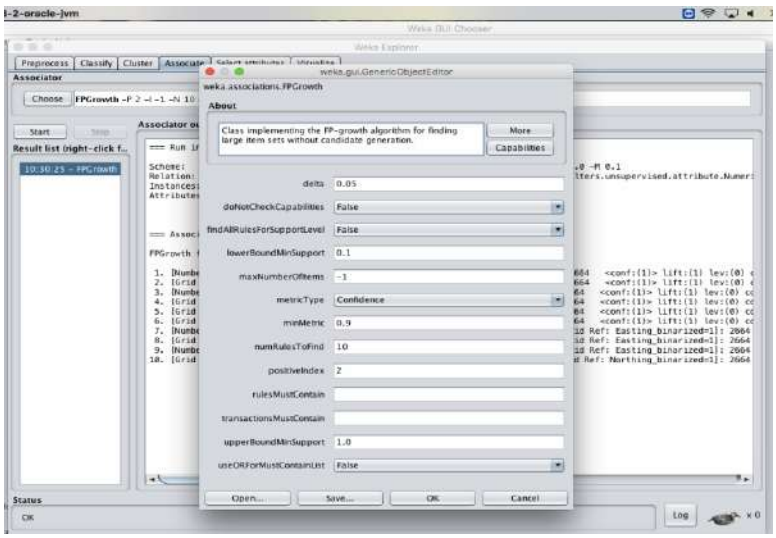
STEPS INVOLVED :

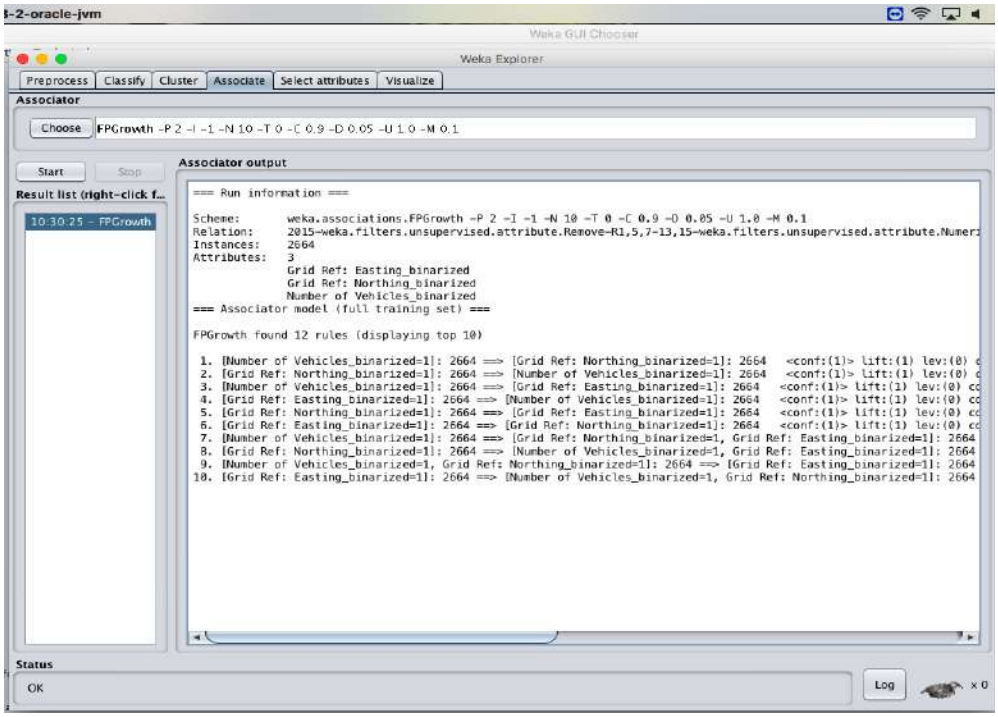
- Choose a set of attributes for clustering and for giving a motivation.
- Choose the dataset and import the dataset into Weka tool.
- Discretize the attributes from all data types to nominal to perform the algorithm.
- Associate the attributes with the FP growth algorithm.

- Set the Upper bound min_sup and lower bound min_sup values.

OBSERVATIONS :

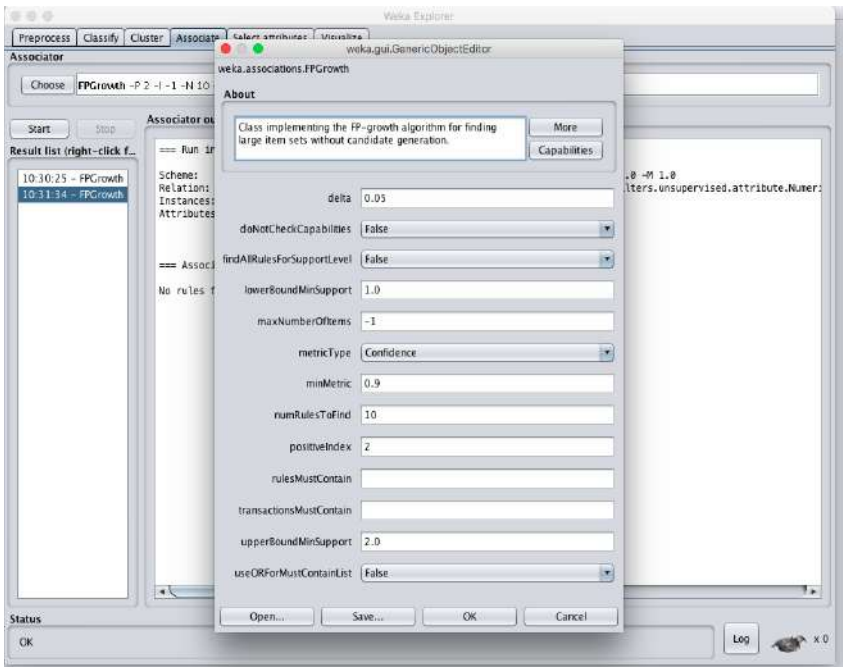
1) When the association rules are of values:
   a) Upper bound min_sup = 1.0
   b) Lower bound min_sup = 0.1
   c) Metric type = confidence.

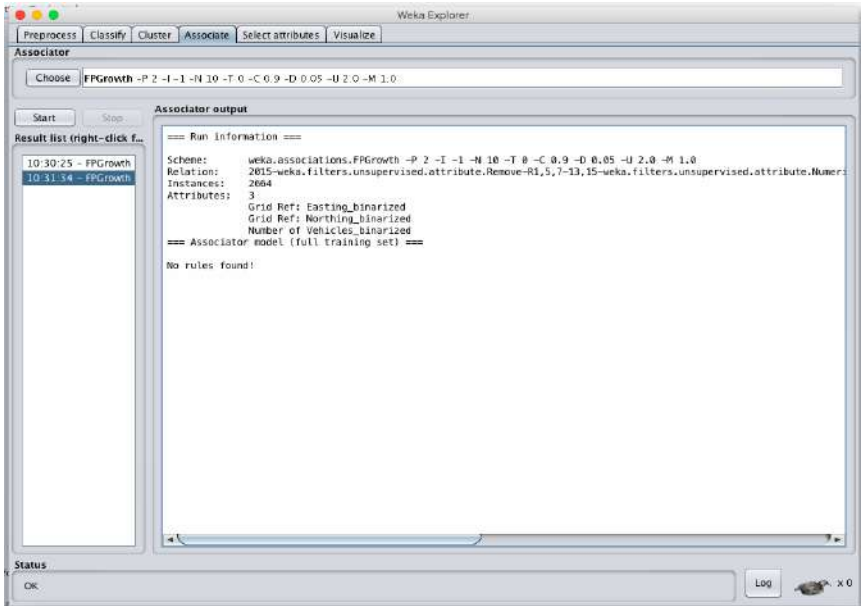




2) When the association rules are of values:

a) Upper bound min_sup = 2.0
b) Lower bound min_sup = 1.0
c) Metric type = confidence.





RESULT :

Thus, the analysis of FP growth algorithm using weka tool has been successfully completed. Incase of changing the upper bound and lower bound values there is a change in the number of rules that are found.

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.

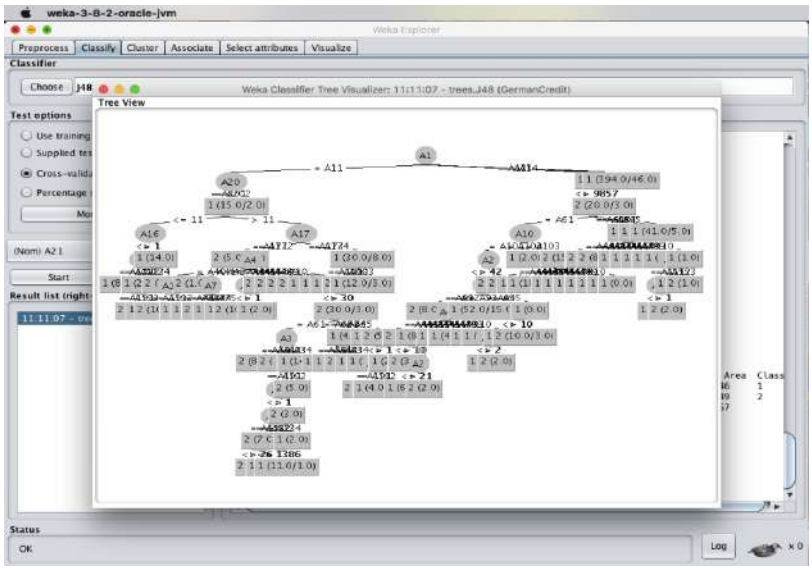
PROCEDURE:

1.Download WEKA And Install
2.Start WEKA

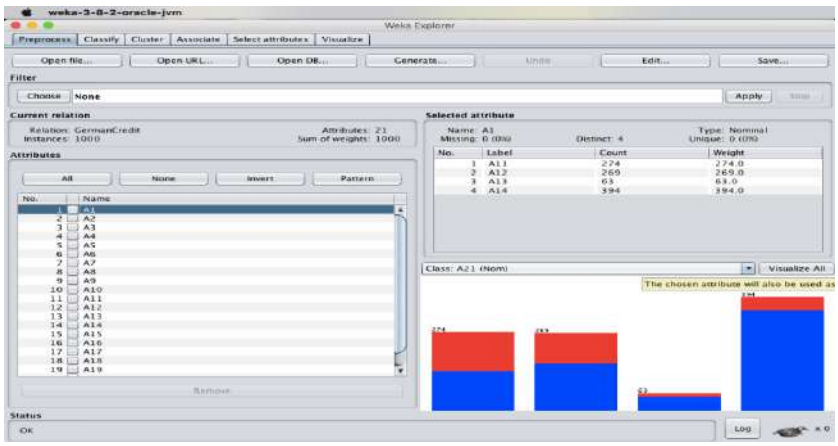
3.Open The Data/iris.arff Dataset
4.Select And Run An Algorithm
5.Review The Results

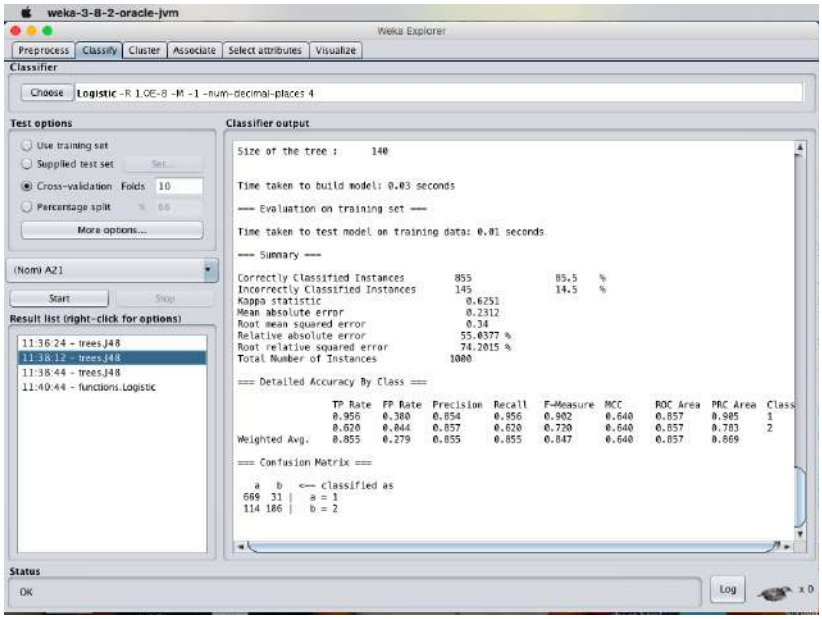
- **Decision Tree :**

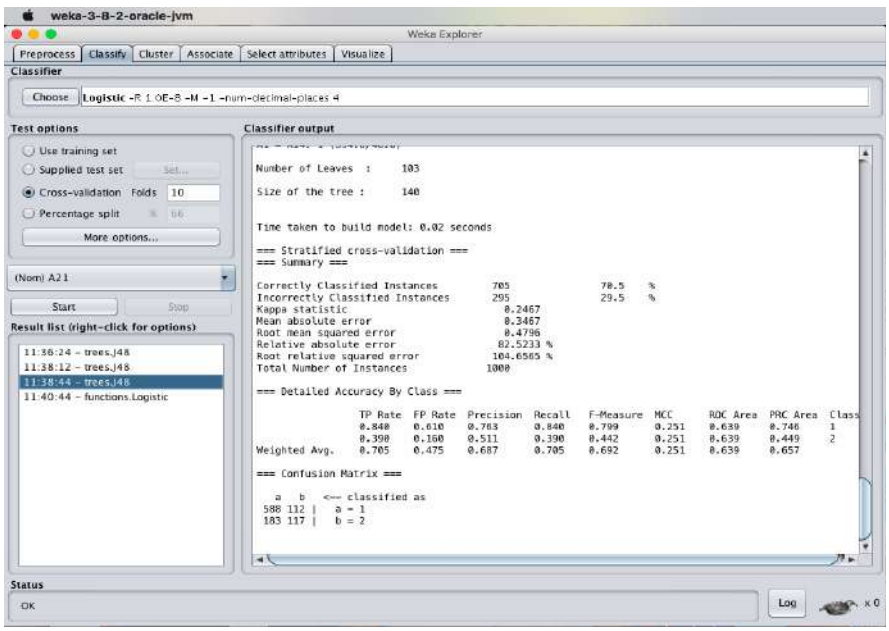
Visualize the decision tree for the given dataset.



or.

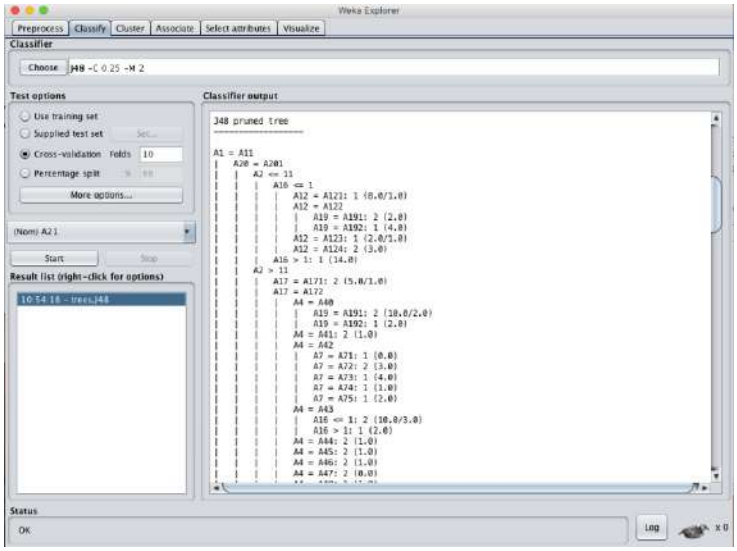


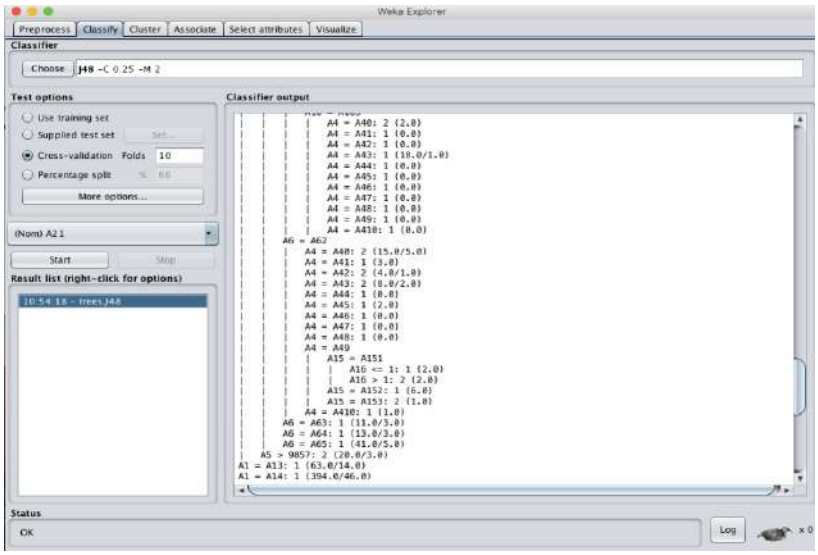


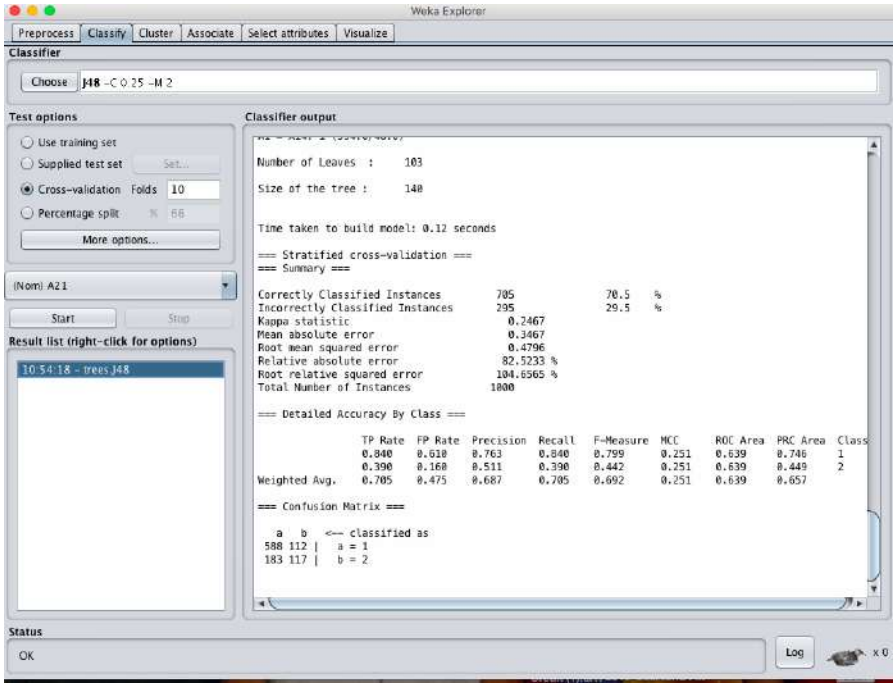


- **CROSS VALIDATION ANALYSIS :**

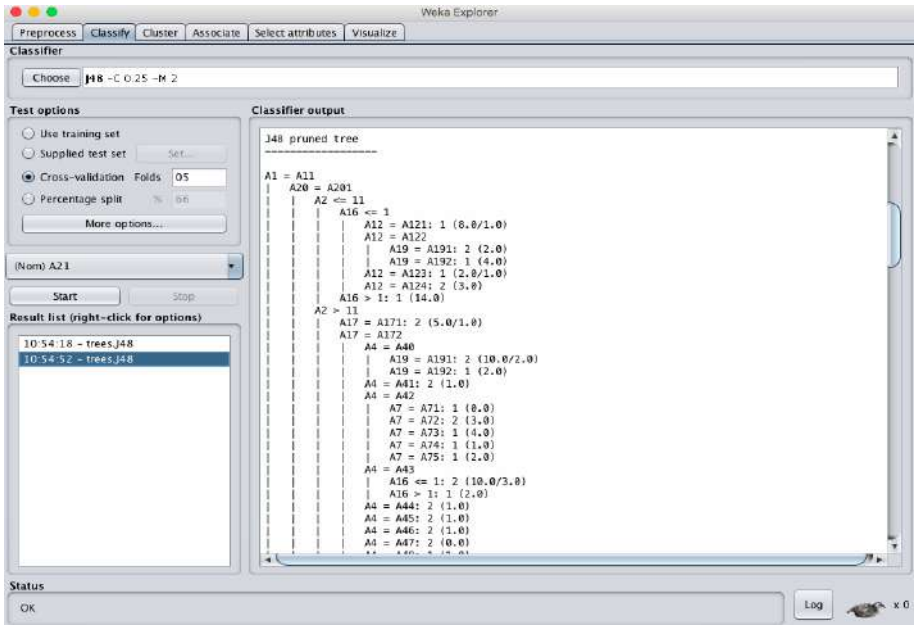
When cross validation folds are 10 :

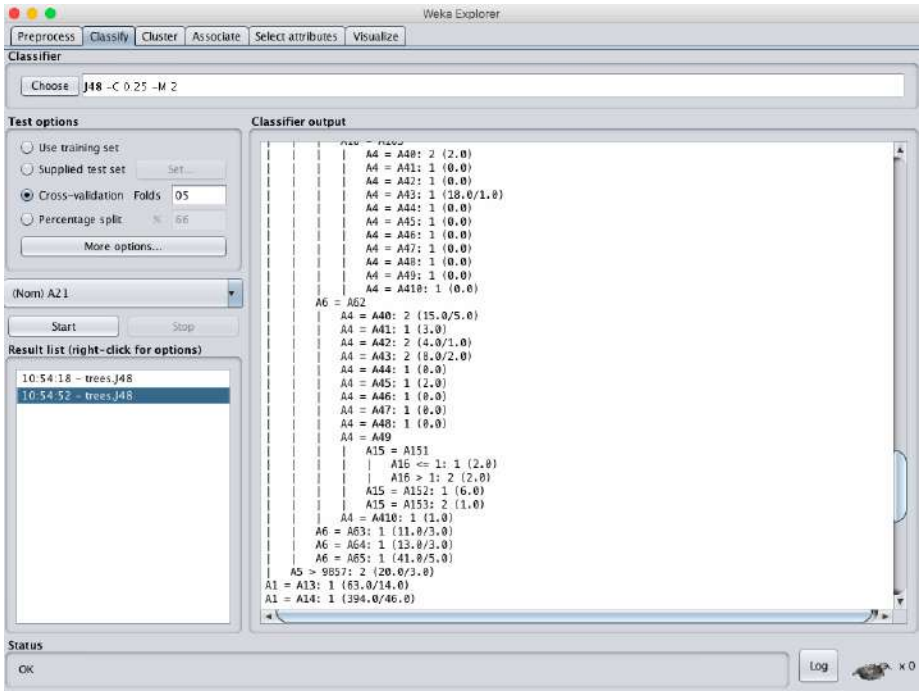

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) A21
Start
Stop
Result list (right-click for options)
10:54:18 - trees.J48
Classifier output
J48 pruned tree
A1 = A11
| A28 = A201
| | A2 <= 11
| | | A16 <= 1
| | | | A12 = A121: 1 (8.0/1.0)
| | | | A12 = A122
| | | | | A19 = A191: 2 (2.0)
| | | | | A19 = A192: 1 (4.0)
| | | | A12 = A123: 1 (2.0/1.0)
| | | | A12 = A124: 2 (3.0)
| | | A16 > 1: 1 (14.0)
| | A2 > 11
| | | A17 = A171: 2 (5.0/1.0)
| | | A17 = A172
| | | | A4 = A40
| | | | | A19 = A191: 2 (10.0/2.0)
| | | | | A19 = A192: 1 (2.0)
| | | | A4 = A41: 2 (1.0)
| | | | A4 = A42
| | | | | A7 = A71: 1 (0.0)
| | | | | A7 = A72: 2 (3.0)
| | | | | A7 = A73: 1 (4.0)
| | | | | A7 = A74: 1 (1.0)
| | | | | A7 = A75: 1 (2.0)
| | | | A4 = A43
| | | | | A16 <= 1: 2 (10.0/3.0)
| | | | | A16 > 1: 1 (2.0)
| | | | A4 = A44: 2 (1.0)
| | | | A4 = A45: 2 (1.0)
| | | | A4 = A46: 2 (1.0)
| | | | A4 = A47: 2 (0.0)
Status
OK
Log
x 0

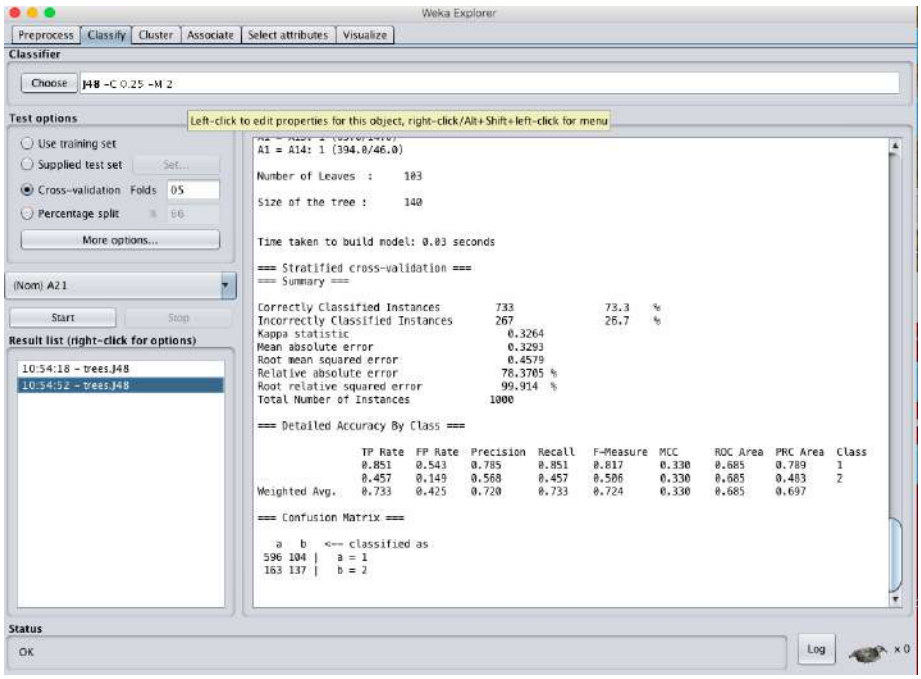

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) A21
Start
Stop
Result list (right-click for options)
10:54:18 - trees.J48
Classifier output
| | | | | A4 = A40: 2 (2.0)
| | | | | A4 = A41: 1 (0.0)
| | | | | A4 = A42: 1 (0.0)
| | | | | A4 = A43: 1 (18.0/1.0)
| | | | | A4 = A44: 1 (0.0)
| | | | | A4 = A45: 1 (0.0)
| | | | | A4 = A46: 1 (0.0)
| | | | | A4 = A47: 1 (0.0)
| | | | | A4 = A48: 1 (0.0)
| | | | | A4 = A49: 1 (0.0)
| | | | | A4 = A410: 1 (0.0)
| | | A6 = A62
| | | | A4 = A40: 2 (15.0/5.0)
| | | | A4 = A41: 1 (3.0)
| | | | A4 = A42: 2 (4.0/1.0)
| | | | A4 = A43: 2 (8.0/2.0)
| | | | A4 = A44: 1 (0.0)
| | | | A4 = A45: 1 (2.0)
| | | | A4 = A46: 1 (0.0)
| | | | A4 = A47: 1 (0.0)
| | | | A4 = A48: 1 (0.0)
| | | | A4 = A49
| | | | | A15 = A151
| | | | | | A16 <= 1: 1 (2.0)
| | | | | | A16 > 1: 2 (2.0)
| | | | | A15 = A152: 1 (6.0)
| | | | | A15 = A153: 2 (1.0)
| | | | A4 = A410: 1 (1.0)
| | | A6 = A63: 1 (11.0/3.0)
| | | A6 = A64: 1 (13.0/3.0)
| | | A6 = A65: 1 (41.0/5.0)
| A5 > 9857: 2 (20.0/3.0)
A1 = A13: 1 (63.0/14.0)
A1 = A14: 1 (394.0/46.0)
Status
OK
Log
x 0



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.

EX.NO:28
Date:

PREDICTION OF CATEGORICAL DATA USING SMO ALGORTIHM THROUGH WEKA

AIM:

To create prediction of categorical data using SMO Algorithm through weka tool.

DESCRIPTION:

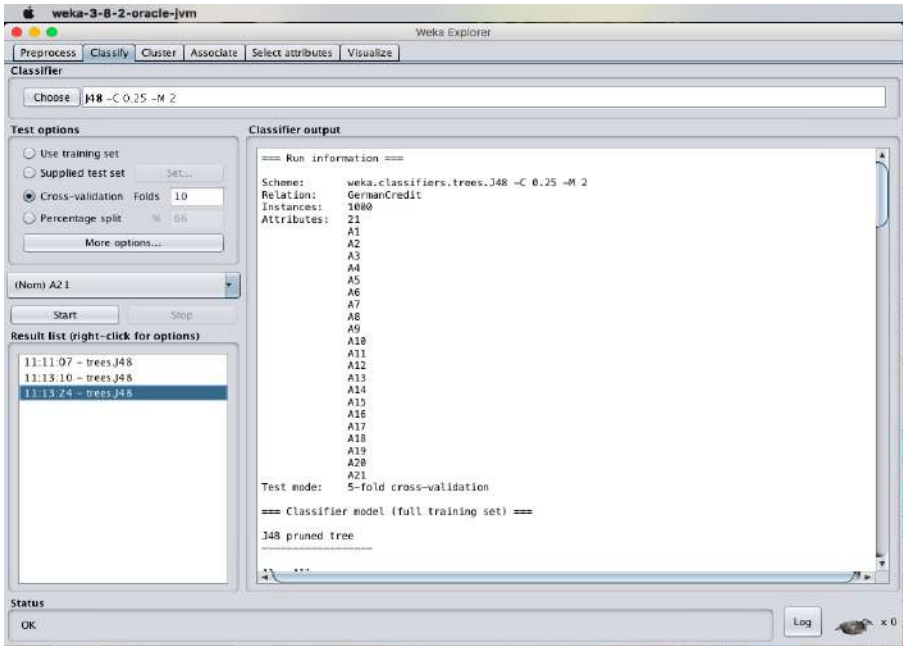
Consider the german credit dataset which can be downloaded from the UCI repository.

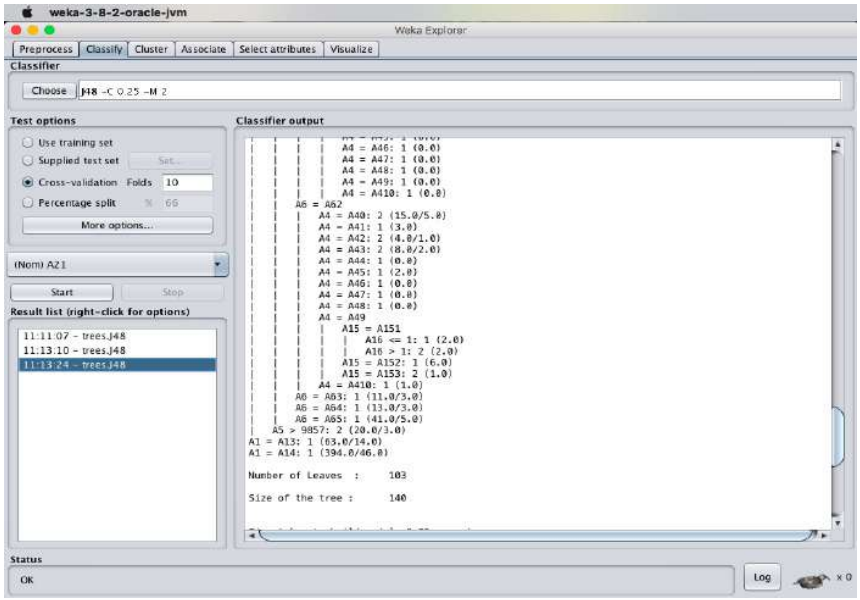
PROCEDURE:

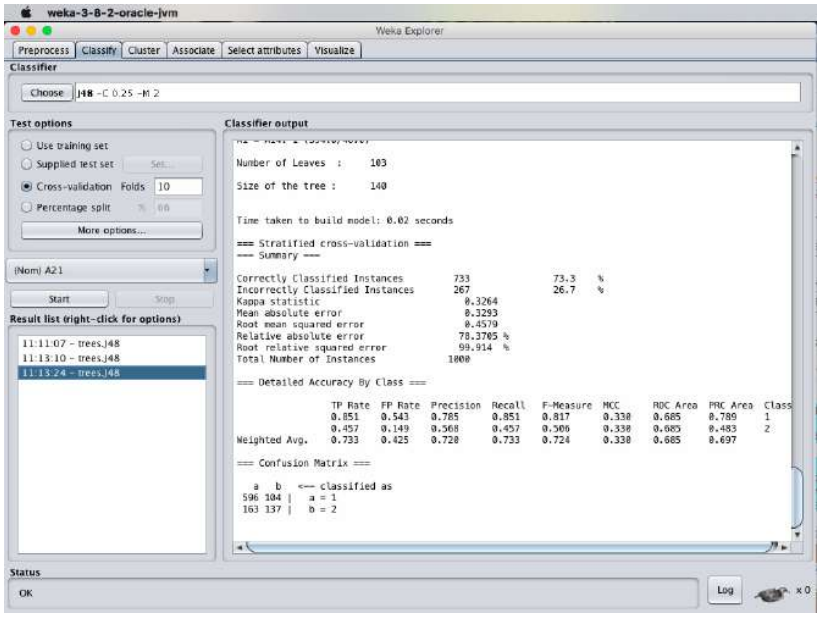
1.Download WEKA And Install
2.Start WEKA
3.Open The Data/iris.arff Dataset
4.Select And Run An Algorithm
5.Review The Results

- **DECISION TREE :**

A tree has many analogies in real life, and turns out that it has influenced a wide area of machine learning, covering both classification and regression. In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. As the name goes, it uses a tree-like model of decisions. Though a commonly used tool in data mining for deriving a strategy to reach a particular goal, its also widely used in machine learning, which will be the main focus of this article.

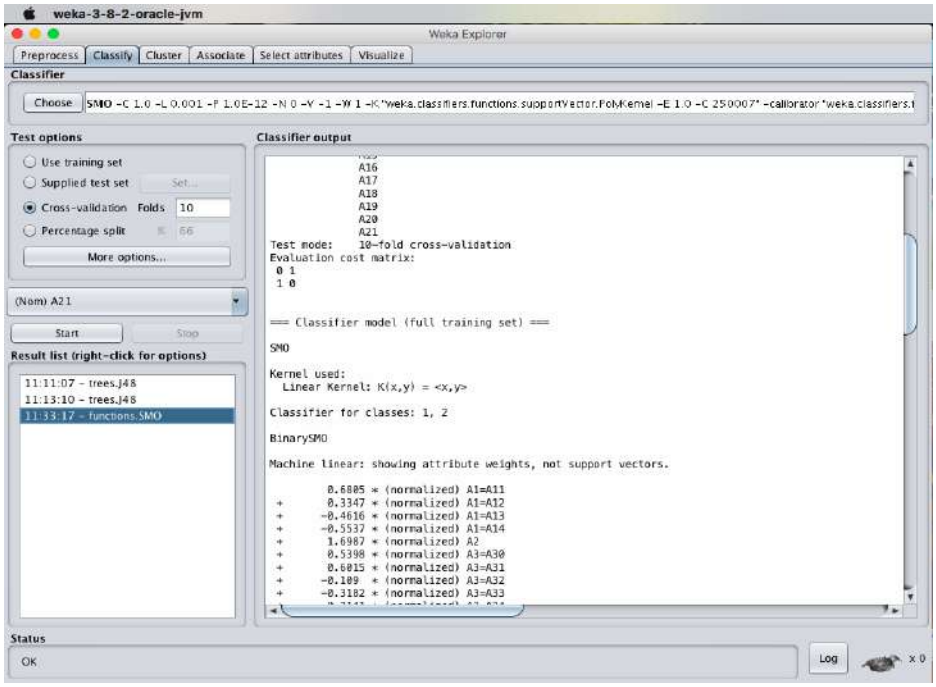

weka-3-8-2-oracle-jvm
Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) A21
Start
Stop
Result list (right-click for options)
11:11:07 - trees.J48
11:13:10 - trees.J48
11:13:24 - trees.J48
Classifier output
=== Run information ===
Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: GermanCredit
Instances: 1000
Attributes: 21
A1
A2
A3
A4
A5
A6
A7
A8
A9
A10
A11
A12
A13
A14
A15
A16
A17
A18
A19
A20
A21
Test mode: 5-fold cross-validation
=== Classifier model (full training set) ===
J48 pruned tree
Status
OK
Log
x 0

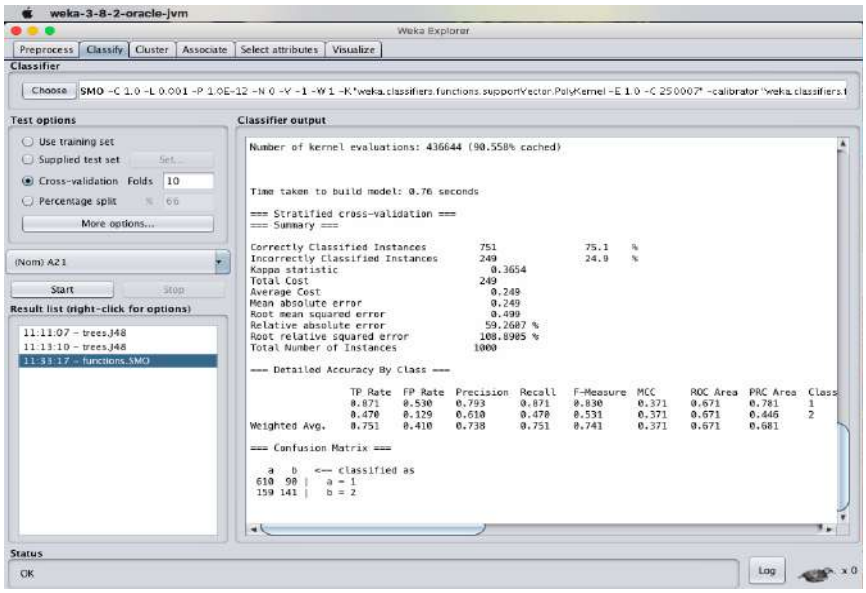

weka-3-8-2-oracle-jvm
Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) A21
Start
Stop
Result list (right-click for options)
11:11:07 - trees.J48
11:13:10 - trees.J48
11:13:24 - trees.J48
Classifier output
| | | | A4 = A46: 1 (0.0)
| | | | A4 = A47: 1 (0.0)
| | | | A4 = A48: 1 (0.0)
| | | | A4 = A49: 1 (0.0)
| | | | A4 = A410: 1 (0.0)
| | A6 = A62
| | | A4 = A40: 2 (15.0/5.0)
| | | A4 = A41: 1 (3.0)
| | | A4 = A42: 2 (4.0/1.0)
| | | A4 = A43: 2 (8.0/2.0)
| | | A4 = A44: 1 (0.0)
| | | A4 = A45: 1 (2.0)
| | | A4 = A46: 1 (0.0)
| | | A4 = A47: 1 (0.0)
| | | A4 = A48: 1 (0.0)
| | | A4 = A49
| | | | A15 = A151
| | | | | A16 <= 1: 1 (2.0)
| | | | | A16 > 1: 2 (2.0)
| | | | A15 = A152: 1 (6.0)
| | | | A15 = A153: 2 (1.0)
| | | A4 = A410: 1 (1.0)
| | A6 = A63: 1 (11.0/3.0)
| | A6 = A64: 1 (13.0/3.0)
| | A6 = A65: 1 (41.0/5.0)
| A5 > 9857: 2 (20.0/3.0)
A1 = A13: 1 (63.0/14.0)
A1 = A14: 1 (394.0/46.0)
Number of Leaves : 103
Size of the tree : 140
Status
OK
Log
x 0



❖ SMO ALGORITHM:

The iterative algorithm Sequential Minimal Optimization (SMO) is used for solving quadratic programming (QP) problems. One example where QP problems are relevant is during the training process of support vector machines (SVM). The SMO algorithm is used to solve in this example a constraint optimization problem. John Platt proposed this algorithm in 1998 and it was successfully used since then. We describe here the basics of the algorithm in the light of big data.





1. Set the cost sensitive evaluation and compare the obtained results.

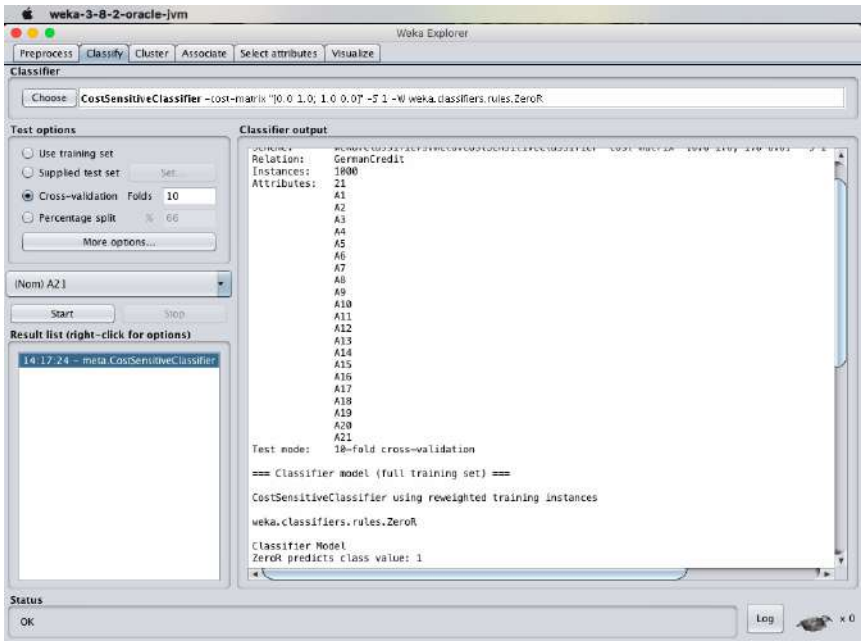
Cost-Sensitive Learning is a type of learning in data mining that takes the misclassification costs (and possibly other types of cost) into consideration. The goal of this type of learning is to minimize the total cost.

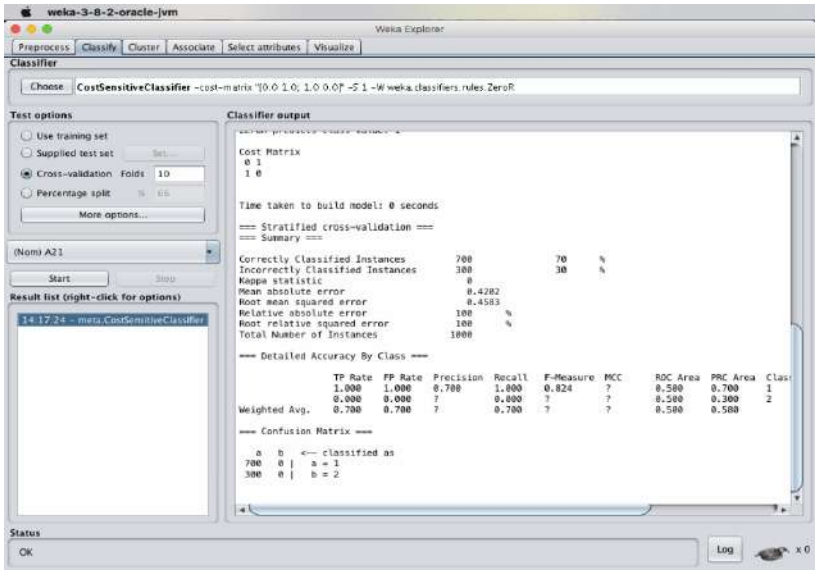
The key difference between cost-sensitive learning and cost-insensitive learning is that cost-sensitive learning treats the different misclassifications differently. Costinsensitive learning does not take the misclassification costs into consideration. The goal of this type of learning is to pursue a high accuracy of classifying examples into a set of known classes.

STEPS :

- Classifythe dataset with the cost sensitive classifier technique.
- Change the cost matrix to 2*2 matrix and execute.

ANALYSIS :





2. What is the significance of the following parameters :

a) Mean Absolute Error :

Mean Absolute Error (MAE) is similar to the Mean Squared Error, but it uses absolute values instead of squaring. This measure is not as popular as MSE, though its meaning is more intuitive (the "average error").

b) Total Number of Instances :

The data present consists of various instances of the class. In the case of german_credit dataset, the total number of instances present in the german credit dataset are 1000 instances.

RESULT :

Thus, the observations and evaluations done on the german_credit dataset are analyzed. The comparison between decision tree and Sequential Minimal Optimization (SMO) has been successfully visualized. In addition to that cost sensitive classifier is been used to analyze few things.

EX.NO:29

Date:

EVALUATING ACCURACY OF THE CLASSIFIERS

AIM:

To create evaluating accuracy of the classifiers using weka tool.

DESCRIPTION:

Consider the german credit dataset which can be downloaded from the UCI repository.

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

ANALYSIS :

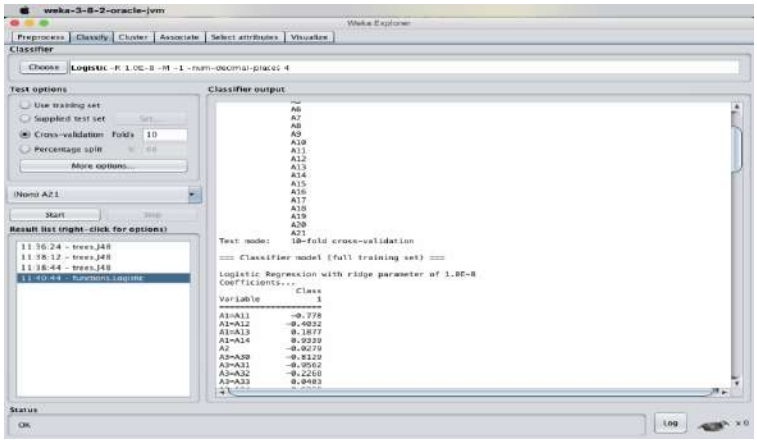
A) Logistic Regression :

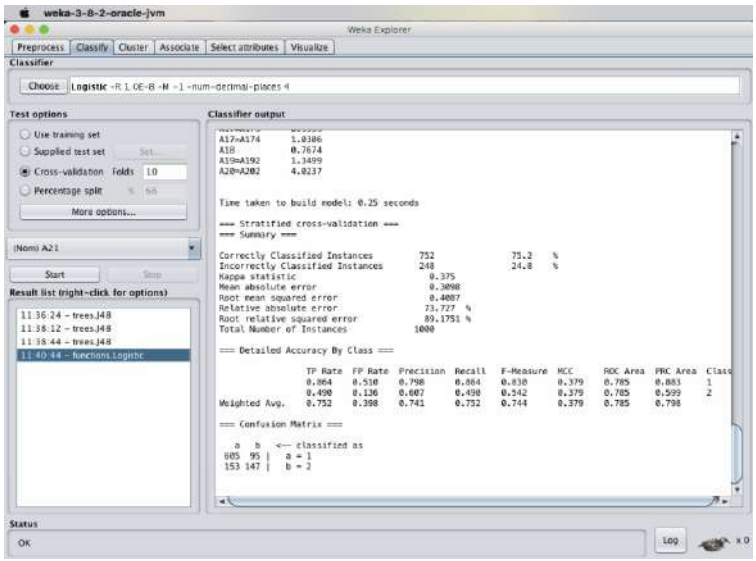
Logistic regression predicts the probability of an outcome that can only have two values (i.e. a dichotomy). The prediction is based on the use of one or several predictors (numerical and categorical).

Steps :

- Load the dataset into the weka tool and preprocess it.
- Apply the classification the logistic regression technique and execute for the result.

Output :





B) Naïve Bayes Algorithm :

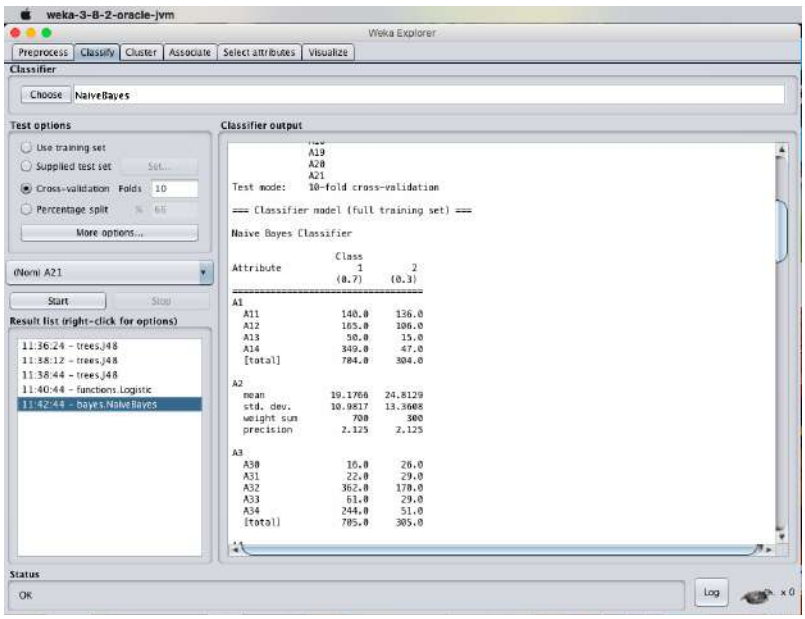
The Naive Bayesian classifier is based on Bayes' theorem with the independence assumptions between predictors. A Naive Bayesian model is easy to build, with no complicated iterative parameter estimation which makes it particularly useful for very large datasets. Despite its simplicity, the

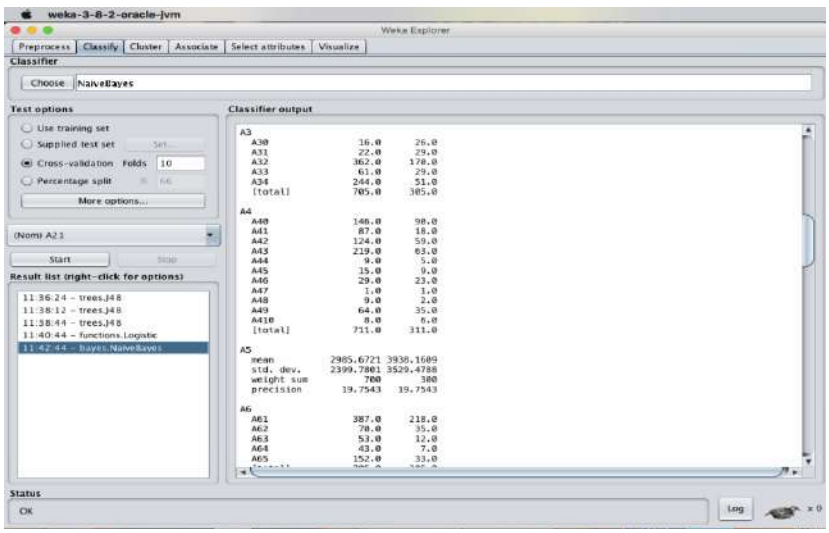
Naive Bayesian classifier often does surprisingly well and is widely used because it often outperforms more sophisticated classification methods.

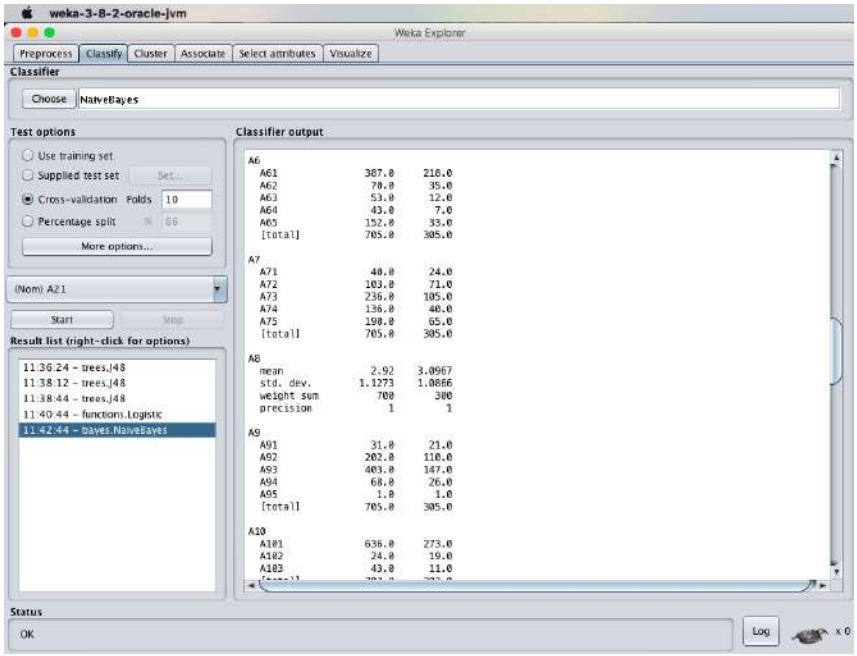
Steps :

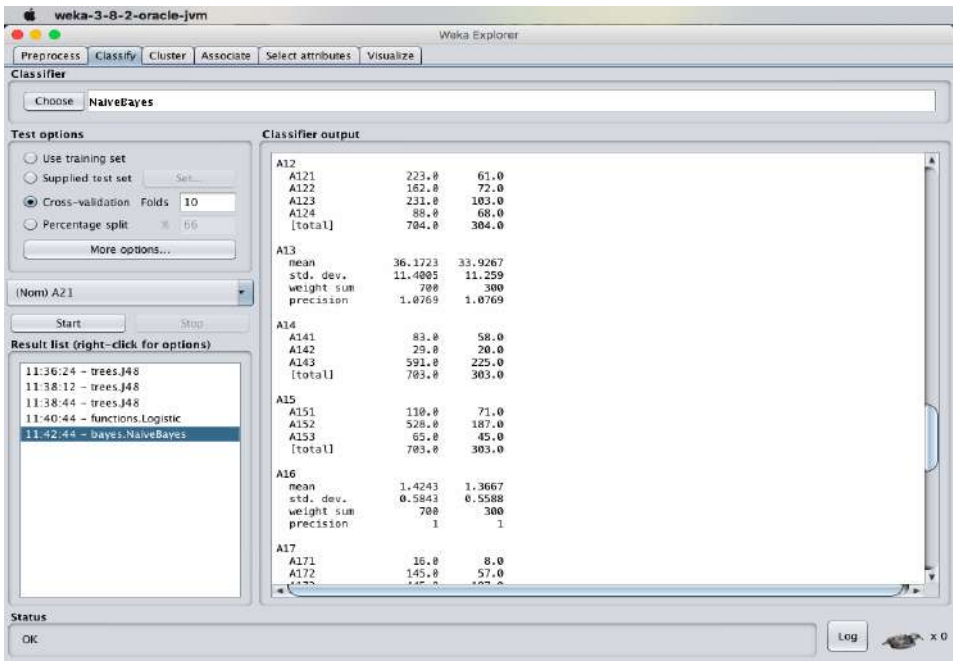
- Load the dataset into the weka tool and preprocess it.
- Apply the classification the Naïve bayes technique and execute for the result.

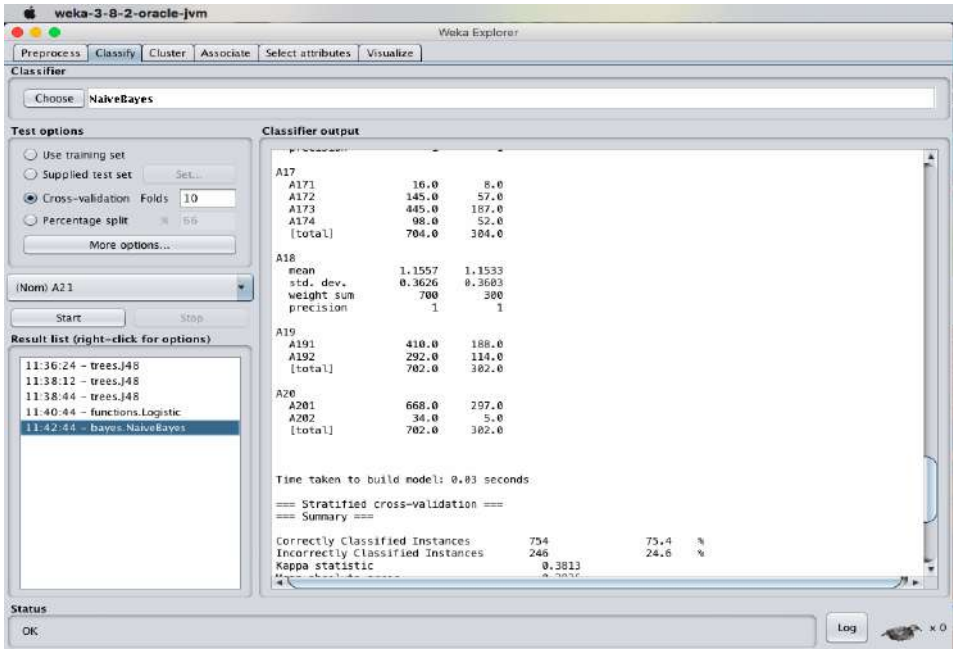
Output :

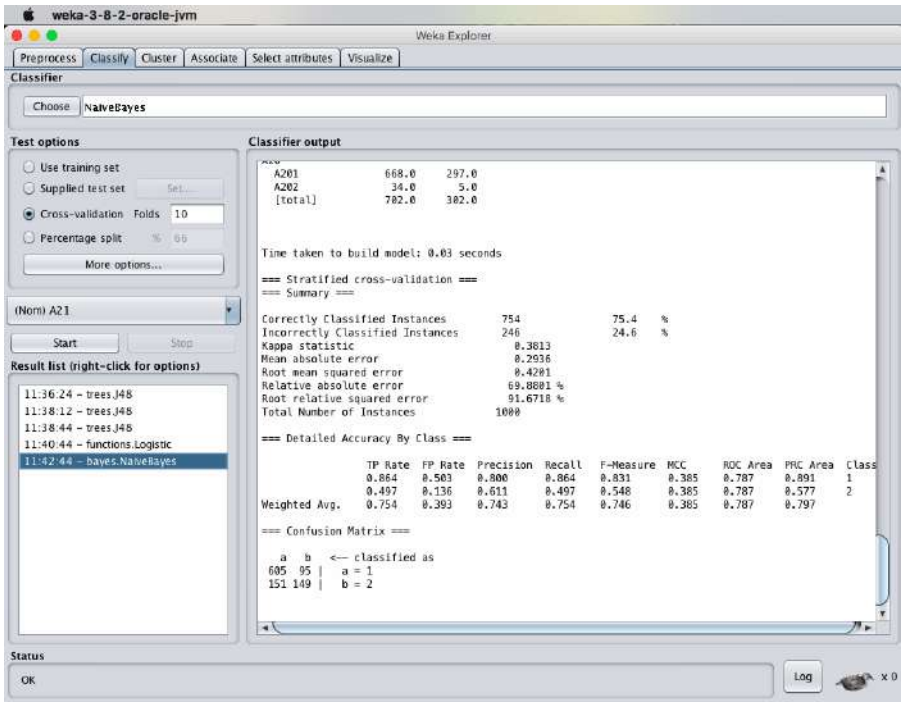





weka-3-8-2-oracle-jvm
Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
NaiveBayes
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) A21
Start
Stop
Result list (right-click for options)
11:36:24 - trees.J48
11:38:12 - trees.J48
11:38:44 - trees.J48
11:40:44 - functions.Logistic
11:42:44 - bayes.NaiveBayes
Classifier output
A6
A61 387.0 218.0
A62 70.0 35.0
A63 53.0 12.0
A64 43.0 7.0
A65 152.0 33.0
[total] 705.0 305.0
A7
A71 40.0 24.0
A72 103.0 71.0
A73 236.0 105.0
A74 136.0 40.0
A75 190.0 65.0
[total] 705.0 305.0
A8
mean 2.92 3.0967
std. dev. 1.1273 1.0866
weight sum 700 300
precision 1 1
A9
A91 31.0 21.0
A92 202.0 110.0
A93 403.0 147.0
A94 68.0 26.0
A95 1.0 1.0
[total] 705.0 305.0
A10
A101 636.0 273.0
A102 24.0 19.0
A103 43.0 11.0
Status
OK
Log
x 0


weka-3-8-2-oracle-jvm
Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
NaiveBayes
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
11:42:44 - bayes.NaiveBayes
Classifier output
A12
A121 223.0 61.0
A122 162.0 72.0
A123 231.0 103.0
A124 88.0 68.0
[total] 704.0 304.0
A13
mean 36.1723 33.9267
std. dev. 11.4005 11.259
weight sum 700 300
precision 1.0769 1.0769
A14
A141 83.0 58.0
A142 29.0 20.0
A143 591.0 225.0
[total] 703.0 303.0
A15
A151 110.0 71.0
A152 528.0 187.0
A153 65.0 45.0
[total] 703.0 303.0
A16
mean 1.4243 1.3667
std. dev. 0.5843 0.5588
weight sum 700 300
precision 1 1
A17
A171 16.0 8.0
A172 145.0 57.0
Status
OK
Log
x 0





C) J48 Algorithm :

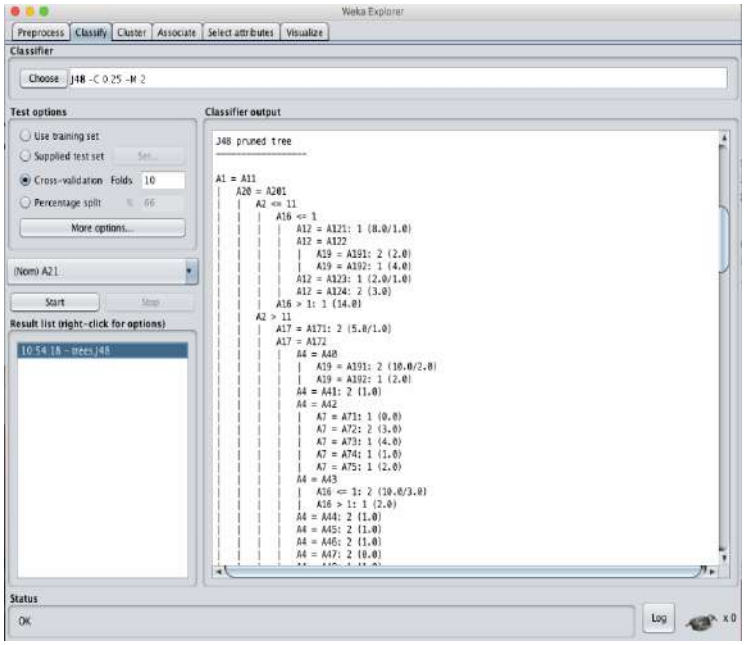
Classification is the process of building a model of classes from a set of records that contain class labels. Decision Tree Algorithm is to find out the way the attributes-vector behaves for a number of instances. Also on the bases of the training instances the classes for the newly generated instances are being found. This algorithm generates the rules for the prediction of the

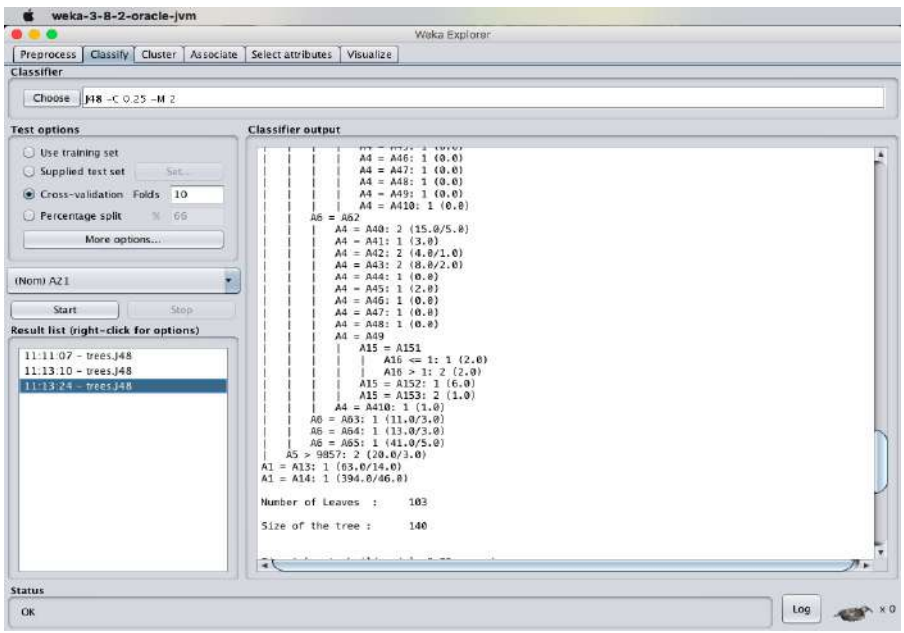
target variable. With the help of tree classification algorithm the critical distribution of the data is easily understandable.

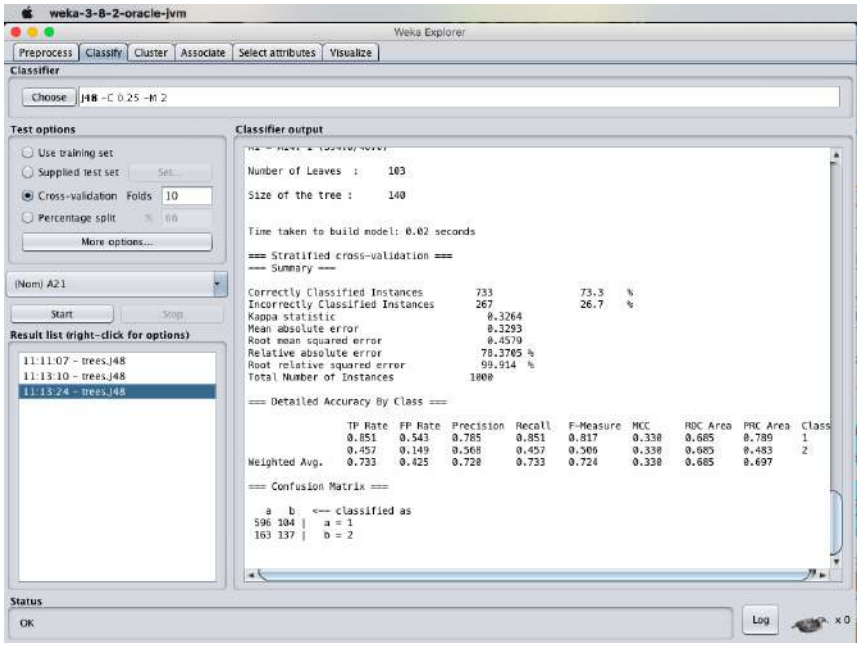
Steps :

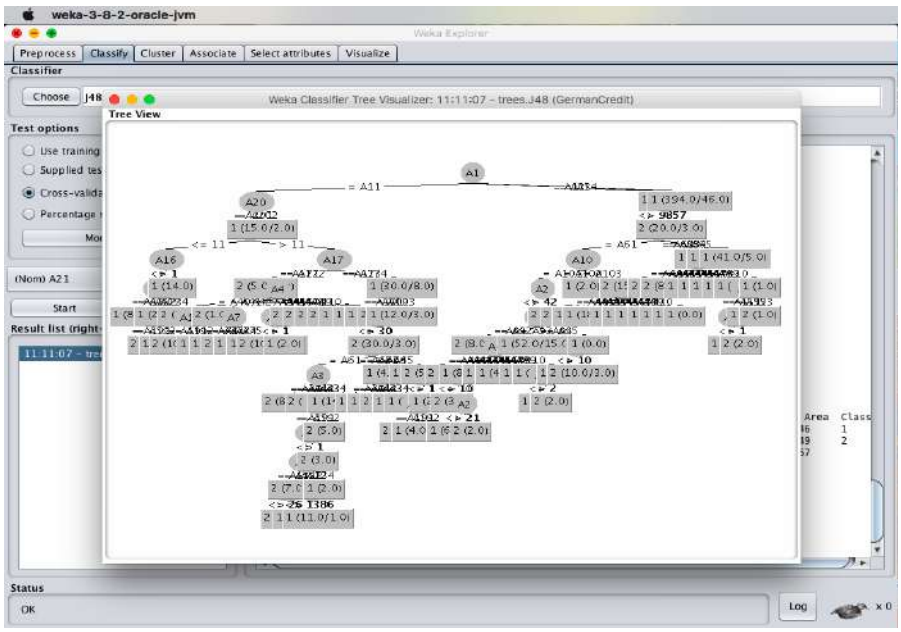
- Load the dataset into the weka tool and preprocess it.
- Apply the classification the J48 technique and execute for the result.

Output :









D) K-Nearest Neighbor :

K-Nearest Neighbors is one of the most basic yet essential classification algorithms in Machine Learning. It belongs to the supervised learning domain and finds intense application in pattern recognition, data mining and intrusion detection.

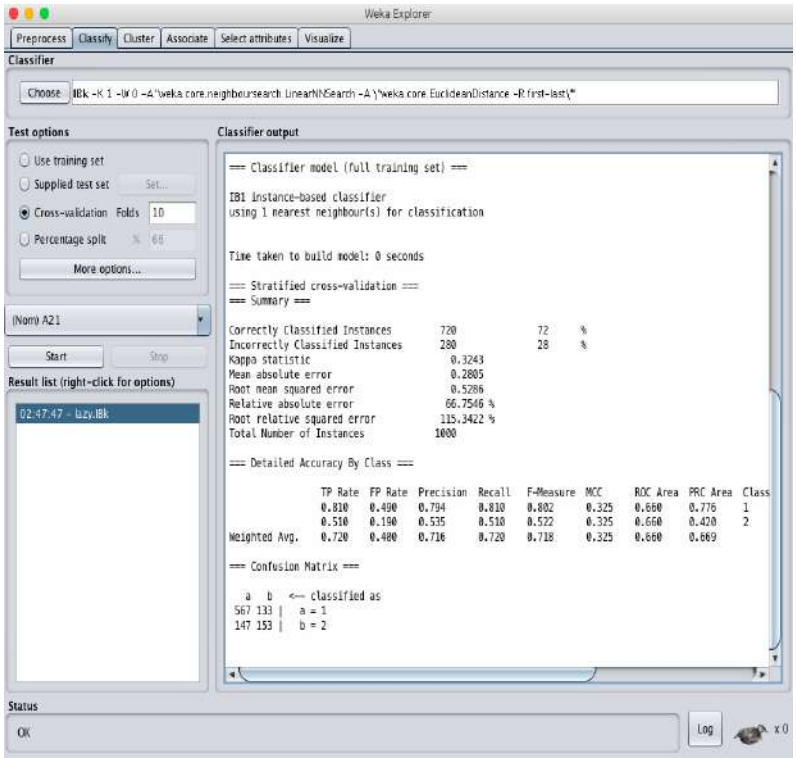
It is widely disposable in real-life scenarios since it is non-parametric, meaning, it does not make any underlying assumptions about the distribution of data (as opposed to other algorithms such as GMM, which assume a Gaussian distribution of the given data).

We are given some prior data (also called training data), which classifies coordinates into groups identified by an attribute.

Steps :

- Load the dataset into the weka tool and preprocess it.
- Apply the classification the K- Nearest Neighbor technique and execute for the result.

Output :

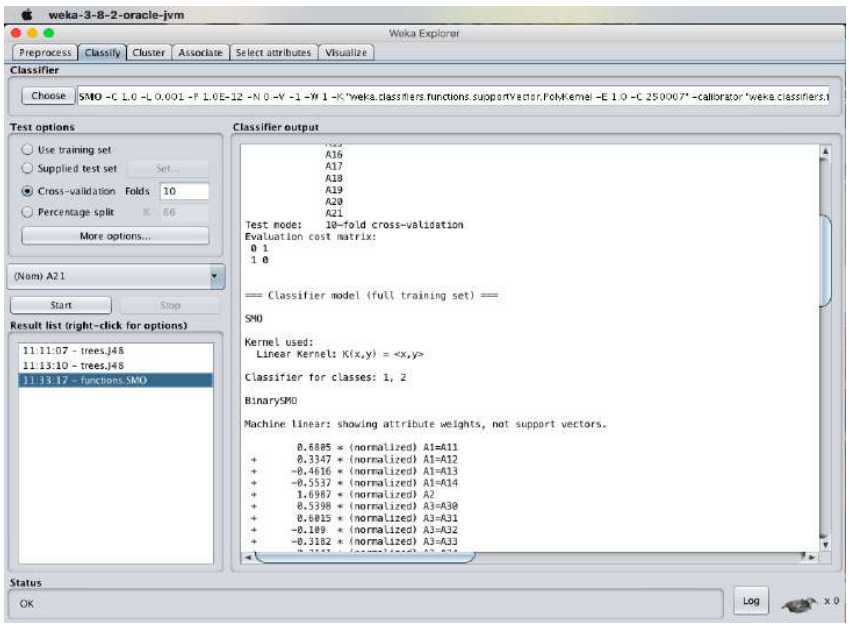


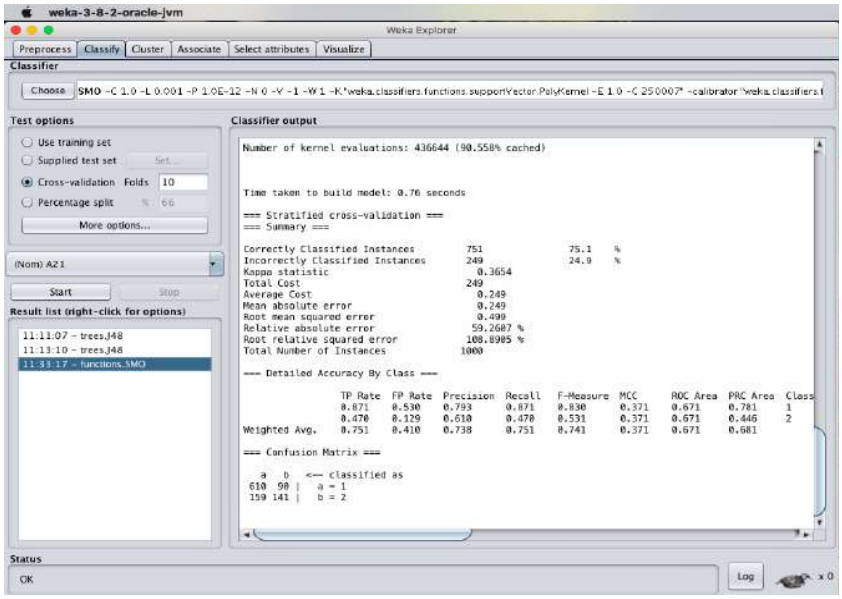
E) SMO Algorithm :

The iterative algorithm Sequential Minimal Optimization (SMO) is used for solving quadratic programming (QP) problems. One example where QP problems are relevant is during the training process of support vector machines (SVM). The SMO algorithm is used to solve in this example a constraint optimization problem. John Platt proposed this algorithm in 1998 and it was successfully used since then. We describe here the basics of the algorithm in the light of big data.

Steps :

- Load the dataset into the weka tool and preprocess it.
- Apply the classification the Sequential Minimal Optimization (SMO)technique and execute for the result.





RESULT :

Thus, the comparison of the confusion matrix for all the methods and techniques. Out of the comparing matrix with all the techniques there is a change in instances. Naïve bayes has more number of correct instances than other but when compared to time K-nearest neighbor is best. The above graphs will show the variations of values in the parameters.

EX.NO:30
Date:

DESCRIPTION NUMERICAL PREDICTION ANALYSIS

USING LINEAR REGRESSION THROUGH WEKA

AIM:

To create description numerical prediction analysis using linear regression through weka tool.

DESCRIPTION:

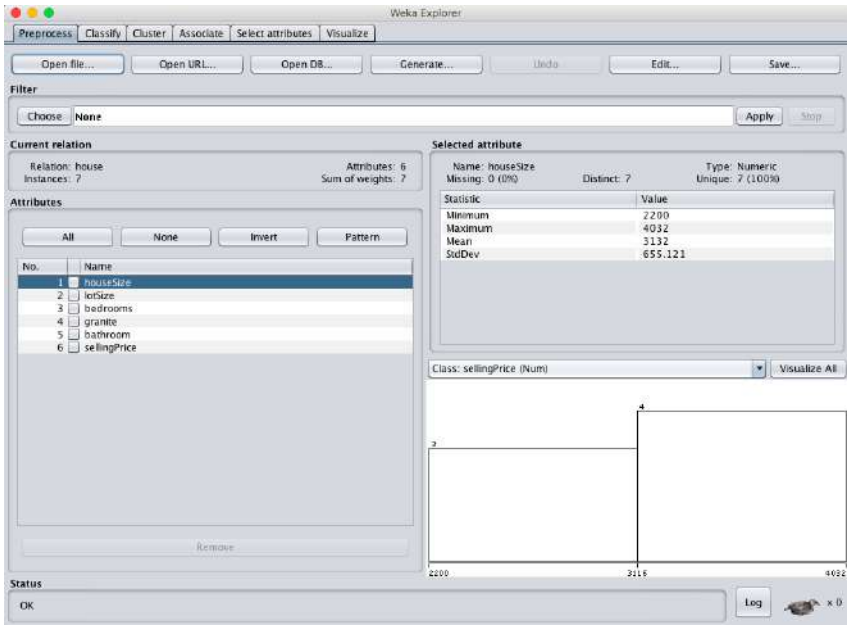
Consider a dataset of house.arff where it contains the attributes as house size, lot size, bedrooms, granite, bathroom and the selling price.

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

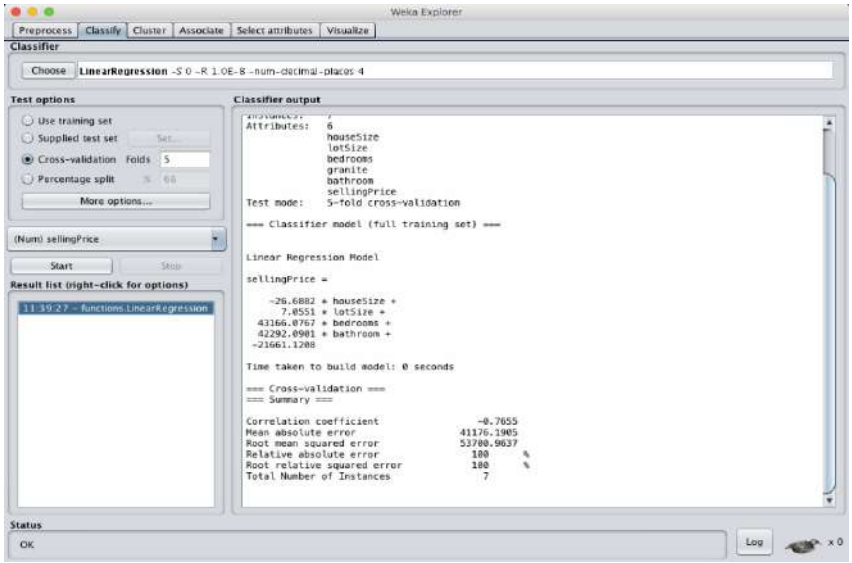
Steps :

- Load the dataset into the weka tool and check for the attributes.
- Classify the data using linear regression analysis method (or) technique.
- Check for the cross-validation folds where the value of the folds should be less than the value of the instances present in the dataset.
- Observe the cross validation summary after applying the linear regression technique for the price of the house.

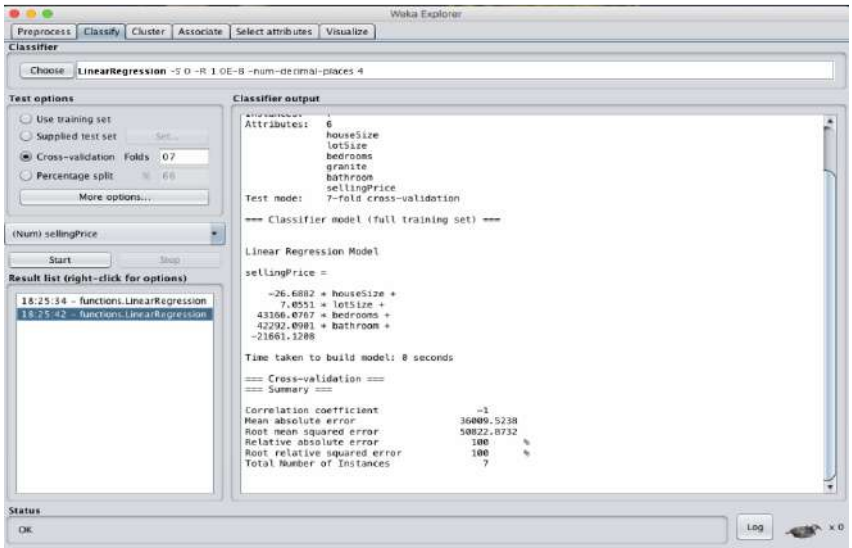


OBSERVATION :

- **When cross validation folds = 05 :**



❖ **When cross validation folds = 10 :**



RESULT :

Thus, the house selling price has been observed using linear regression model. If the value of cross validation folds decreases time for creating model will be less than when folds value high, and the mean absolute error and Root mean square error values decreases with increase in the cross validation folds value.