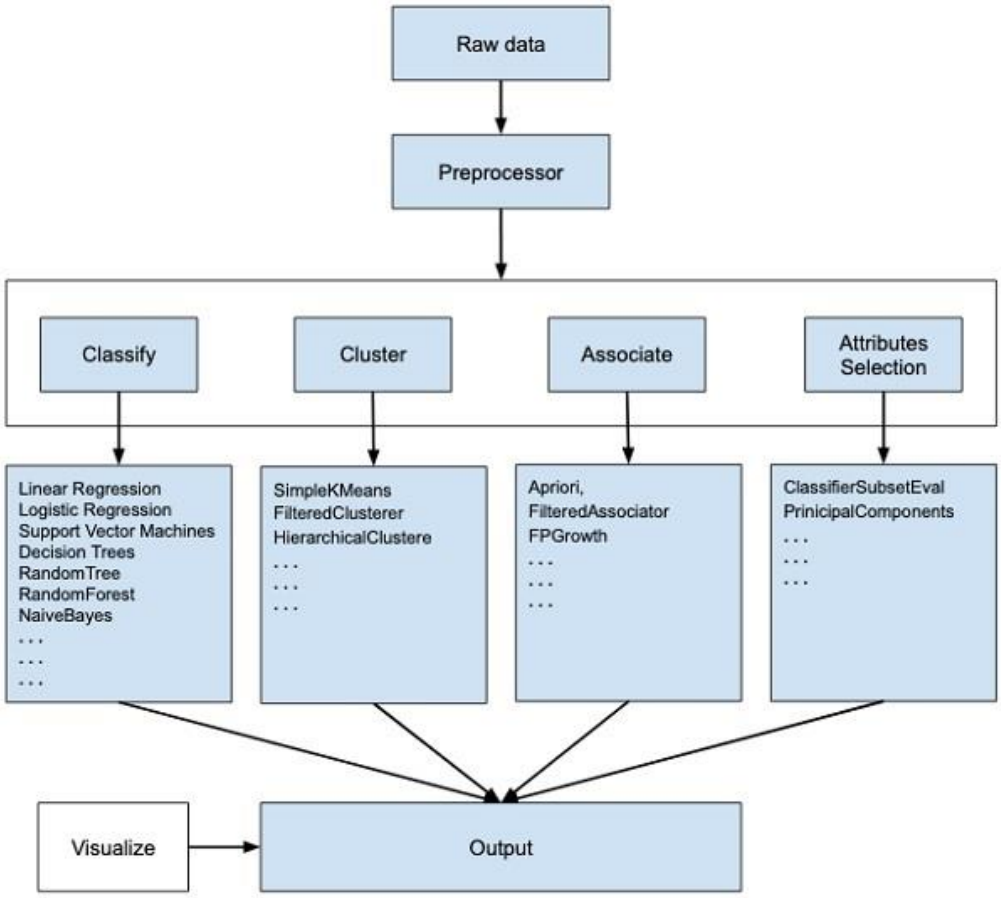
Semester	T.E. Semester VI – Computer Engineering
Subject	Data Warehousing and Mining
Subject Professor In-charge	Prof. Kavita Shirsat
Assisting Teachers	Prof. Kavita Shirsat
Laboratory	M-312B

Student Name	Vibodh Bhosure	
Roll Number	20102A0032	
Grade and Subject Teacher's Signature		

Experiment Number	07	
Experiment Title	Perform data pre-processing task and demonstrate Classification and Clustering on data set using data mining tool (Weka)	
Resources / Apparatus Required	Hardware: Computer system	Software: Weka
Description	What is Weka tool ? WEKA - an open source software provides tools for data preprocessing, implementation of several Machine Learning algorithms, and visualization tools so that you can develop machine learning techniques and apply them to real-world data mining problems. What WEKA offers is summarized in the following diagram −	

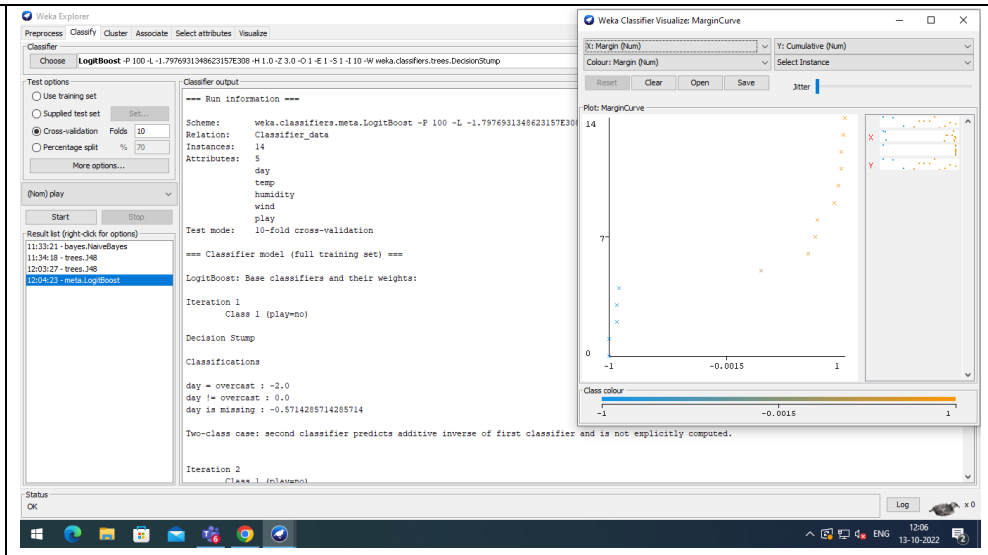


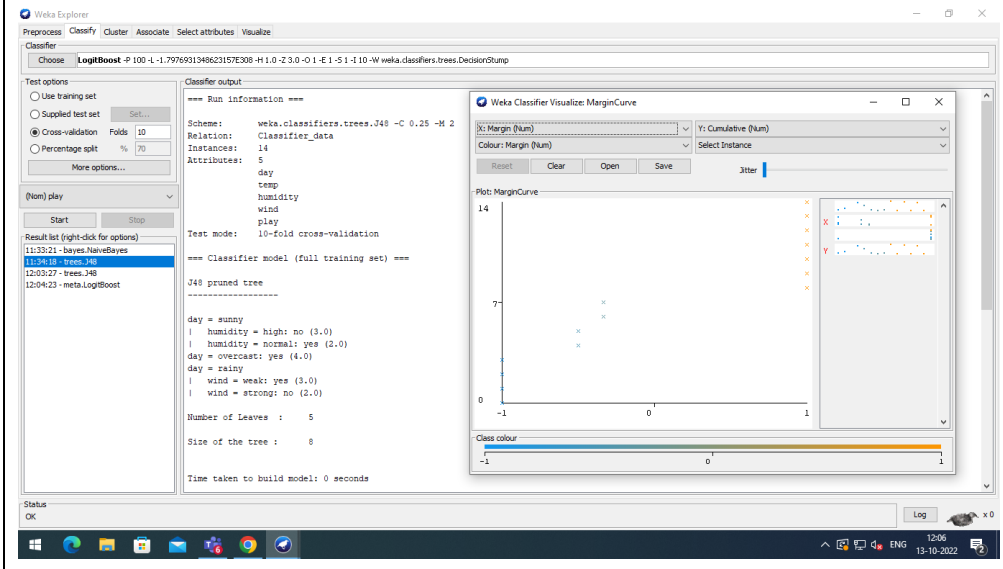
Classification :-
Classification of the data mining system helps users to understand the system and match their requirements with such systems.

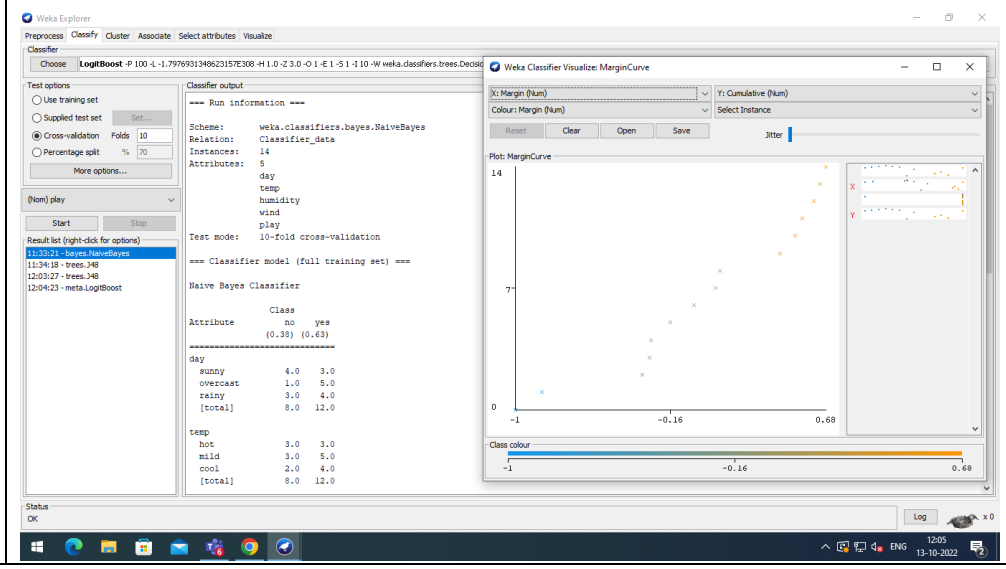
Clustering :-
The process of making a group of abstract objects into classes of similar objects is known as clustering.

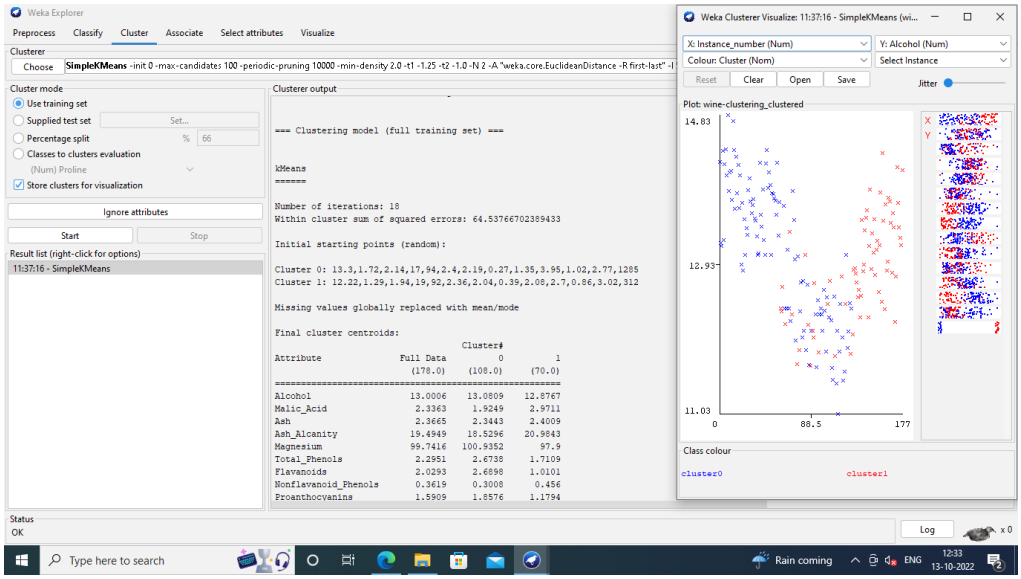
Associate :-
Association Mining searches for frequent items in the data set. In frequent mining usually the interesting associations and correlations between item sets in transactional and relational databases are found. In short, Frequent Mining shows which items appear together in a transaction or relation.

Implementation :-	Classification -


Weka Explorer
Preprocess Classify Cluster Associate Select attributes Visualize
Classifier
Choose LogitBoost -P 100 -L -1.7976931348623157E308 -H 1.0 -Z 3.0 -O 1 -E 1 -S 1 -I 10 -W weka.classifiers.trees.DecisionStump
Test options
Use training set
Supplied test set Set...
Cross-validation Folds 10
Percentage split % 70
More options...
(Nom) play
Start Stop
Result list (right-click for options)
11:33:21 - bayes.NaiveBayes
11:34:18 - trees.J48
12:03:27 - trees.J48
12:04:23 - meta.LogitBoost
Classifier output
=== Run information ===
Scheme: weka.classifiers.meta.LogitBoost -P 100 -L -1.7976931348623157E30
Relation: Classifier_data
Instances: 14
Attributes: 5
day
temp
humidity
wind
play
Test mode: 10-fold cross-validation
=== Classifier model (full training set) ===
LogitBoost: Base classifiers and their weights:
Iteration 1
Class 1 (play=no)
Decision Stump
Classifications
day = overcast : -2.0
day != overcast : 0.0
day is missing : -0.5714285714285714
Two-class case: second classifier predicts additive inverse of first classifier and is not explicitly computed.
Iteration 2
Weka Classifier Visualize: MarginCurve
X: Margin (Num)
Y: Cumulative (Num)
Colour: Margin (Num)
Select Instance
Reset Clear Open Save
Jitter
Plot: MarginCurve
14
7
0
-1
-0.0015
1
Class colour
Status
OK
Log
x 0
ENG
12:06
13-10-2022


Weka Explorer
Preprocess Classify Cluster Associate Select attributes Visualize
Classifier
Choose LogitBoost -P 100 -L -1.7976931348623157E308 -H 1.0 -Z 3.0 -O 1 -E 1 -S 1 -I 10 -W weka.classifiers.trees.DecisionStump
Test options
Use training set
Supplied test set Set...
Cross-validation Folds 10
Percentage split % 70
More options...
(Nom) play
Start Stop
Result list (right-click for options)
11:33:21 - bayes.NaiveBayes
11:34:18 - trees.J48
12:03:27 - trees.J48
12:04:23 - meta.LogitBoost
Classifier output
=== Run information ===
Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: Classifier_data
Instances: 14
Attributes: 5
day
temp
humidity
wind
play
Test mode: 10-fold cross-validation
=== Classifier model (full training set) ===
J48 pruned tree
day = sunny
| humidity = high: no (3.0)
| humidity = normal: yes (2.0)
day = overcast: yes (4.0)
day = rainy
| wind = weak: yes (3.0)
| wind = strong: no (2.0)
Number of Leaves : 5
Size of the tree : 8
Time taken to build model: 0 seconds
Weka Classifier Visualize: MarginCurve
X: Margin (Num)
Y: Cumulative (Num)
Colour: Margin (Num)
Select Instance
Reset Clear Open Save
Jitter
Plot: MarginCurve
14
7
0
-1
0
1
Class colour
Status
OK
Log
x 0
ENG
12:06
13-10-2022


Weka Explorer
Preprocess Classify Cluster Associate Select attributes Visualize
Classifier
Choose LogitBoost -P 100 -L -1.7976931348623157E308 -H 1.0 -Z 3.0 -O 1 -E 1 -S 1 -I 10 -W weka.classifiers.trees.Decisio
Test options
Use training set
Supplied test set Set...
Cross-validation Folds 10
Percentage split % 70
More options...
(Nom) play
Start Stop
Result list (right-click for options)
11:33:21 - bayes.NaiveBayes
11:34:18 - trees.J48
12:03:27 - trees.J48
12:04:23 - meta.LogitBoost
Classifier output
=== Run information ===
Scheme: weka.classifiers.bayes.NaiveBayes
Relation: Classifier_data
Instances: 14
Attributes: 5
day
temp
humidity
wind
play
Test mode: 10-fold cross-validation
=== Classifier model (full training set) ===
Naive Bayes Classifier
Class
Attribute no yes
(0.38) (0.63)
day
sunny 4.0 3.0
overcast 1.0 5.0
rainy 3.0 4.0
[total] 8.0 12.0
temp
hot 3.0 3.0
mild 3.0 5.0
cool 2.0 4.0
[total] 8.0 12.0
Weka Classifier Visualize: MarginCurve
X: Margin (Num)
Y: Cumulative (Num)
Colour: Margin (Num)
Select Instance
Reset Clear Open Save
Jitter
Plot: MarginCurve
14
7
0
-1
-0.16
0.68
Class colour
Status
OK
Log
x 0
ENG
12:05
13-10-2022

	Clustering :- 
Conclusion :-	Here we implemented datamining task of classification and clustering on the weka tool over the data set provided by us and obtained visual plots