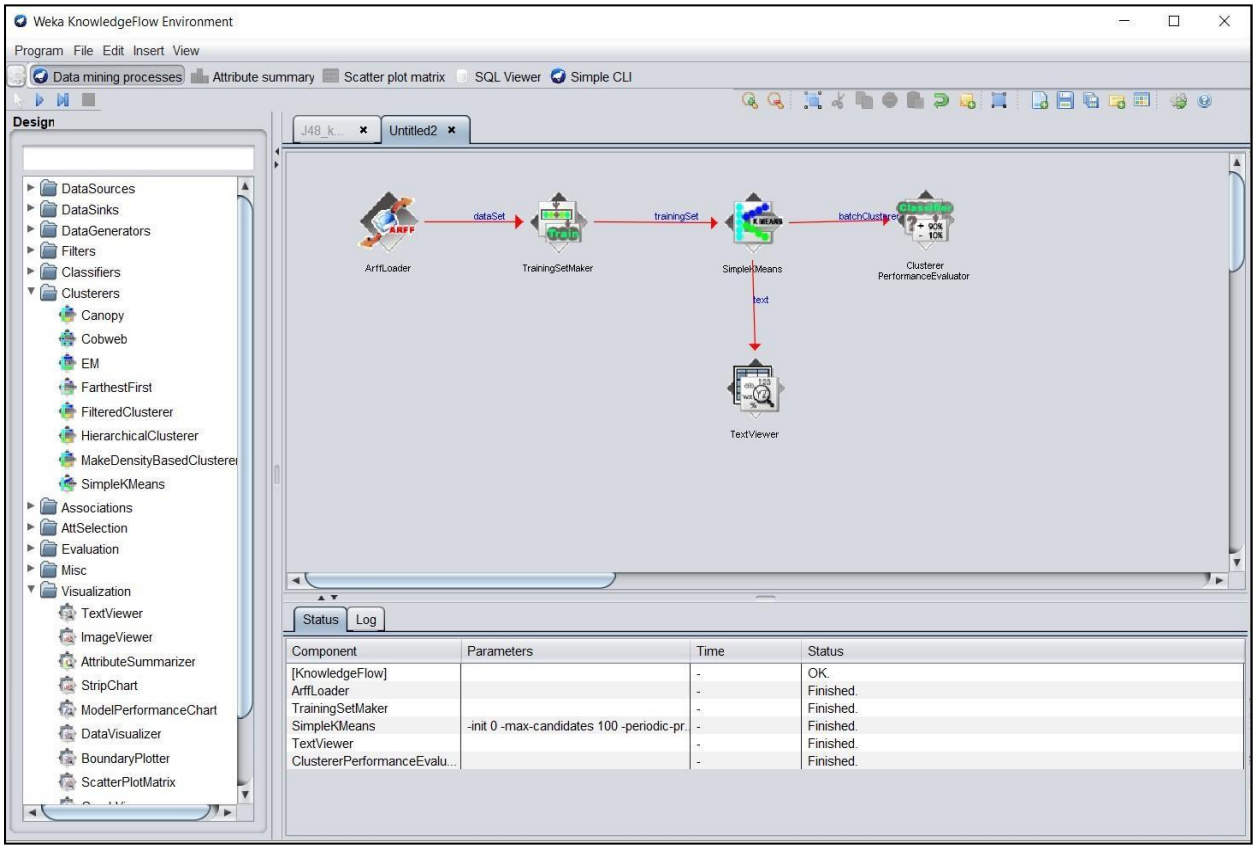
Experiment 7A: K-Means

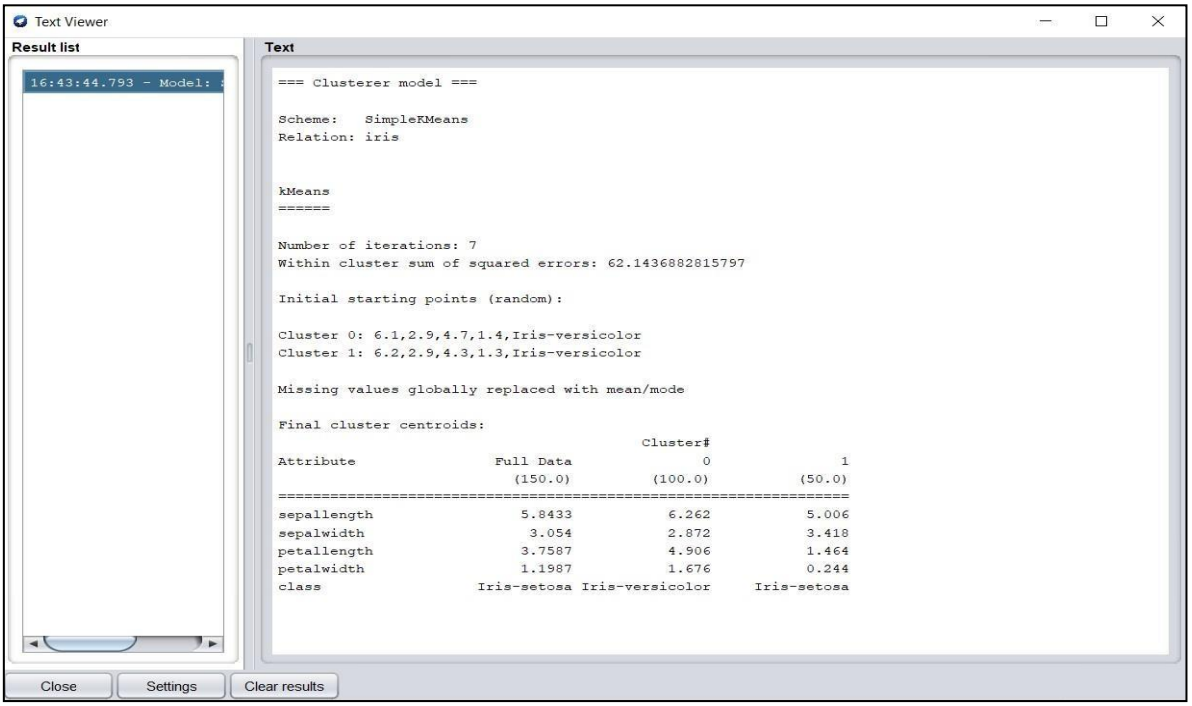
Aim: Perform data Pre-processing task and Demonstrate performing Classification, Clustering, Association algorithm on data sets using data mining tool (WEKA,R tool, XL Miner, etc.)

Knowledge Flow Model:

Designing of K Means Model and Passing the dataset:

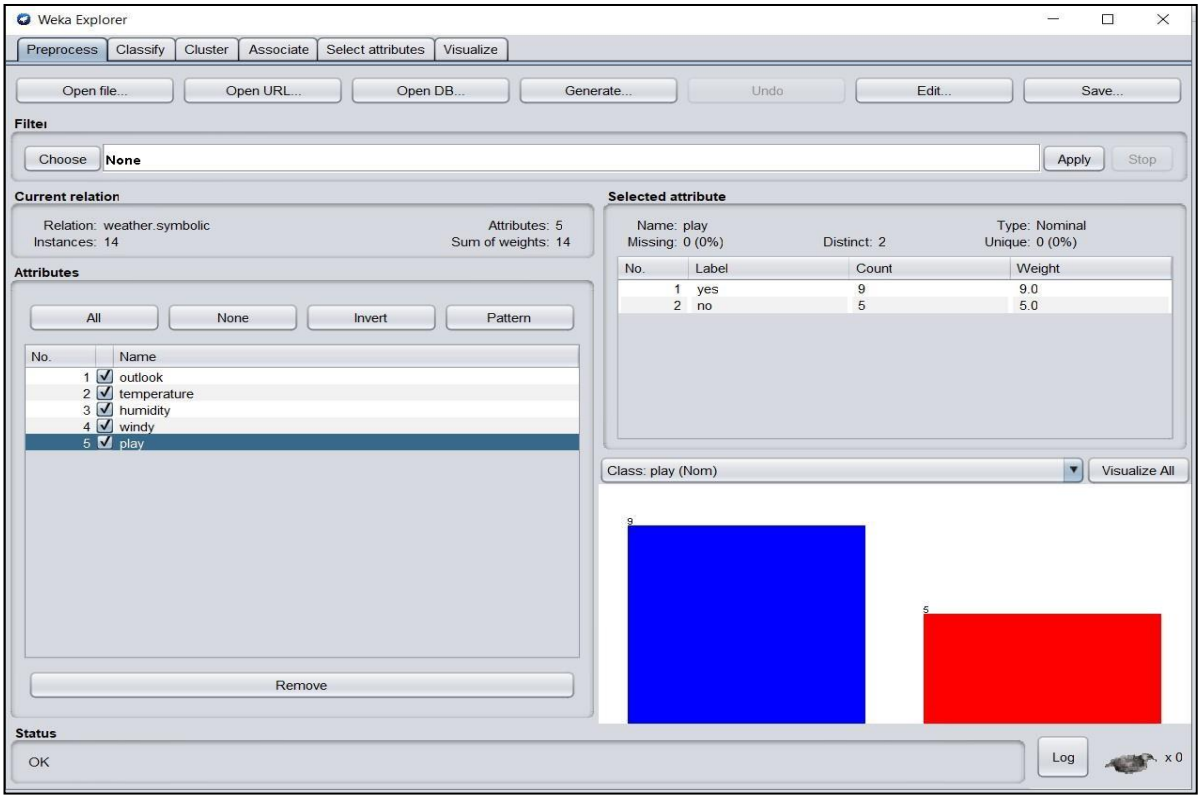


Output of K Means Clustering in TextViewer:

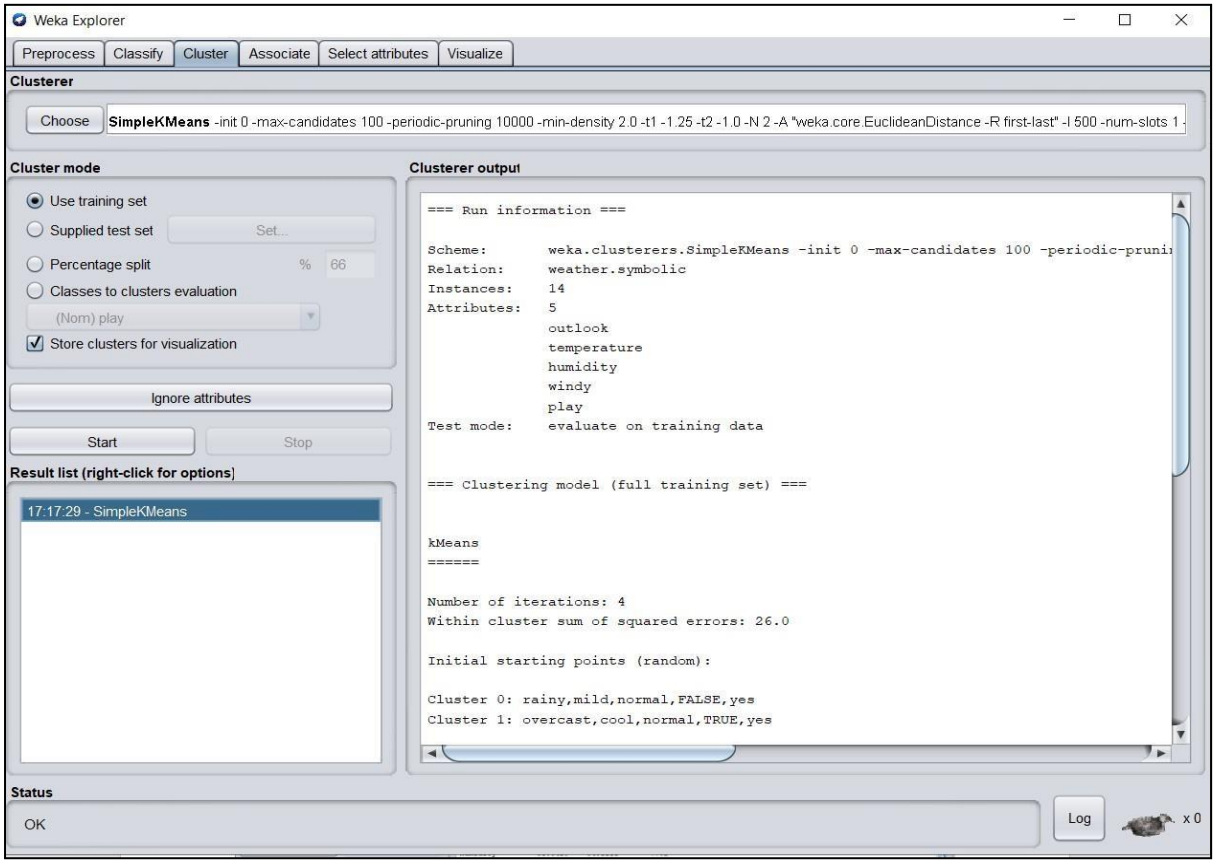


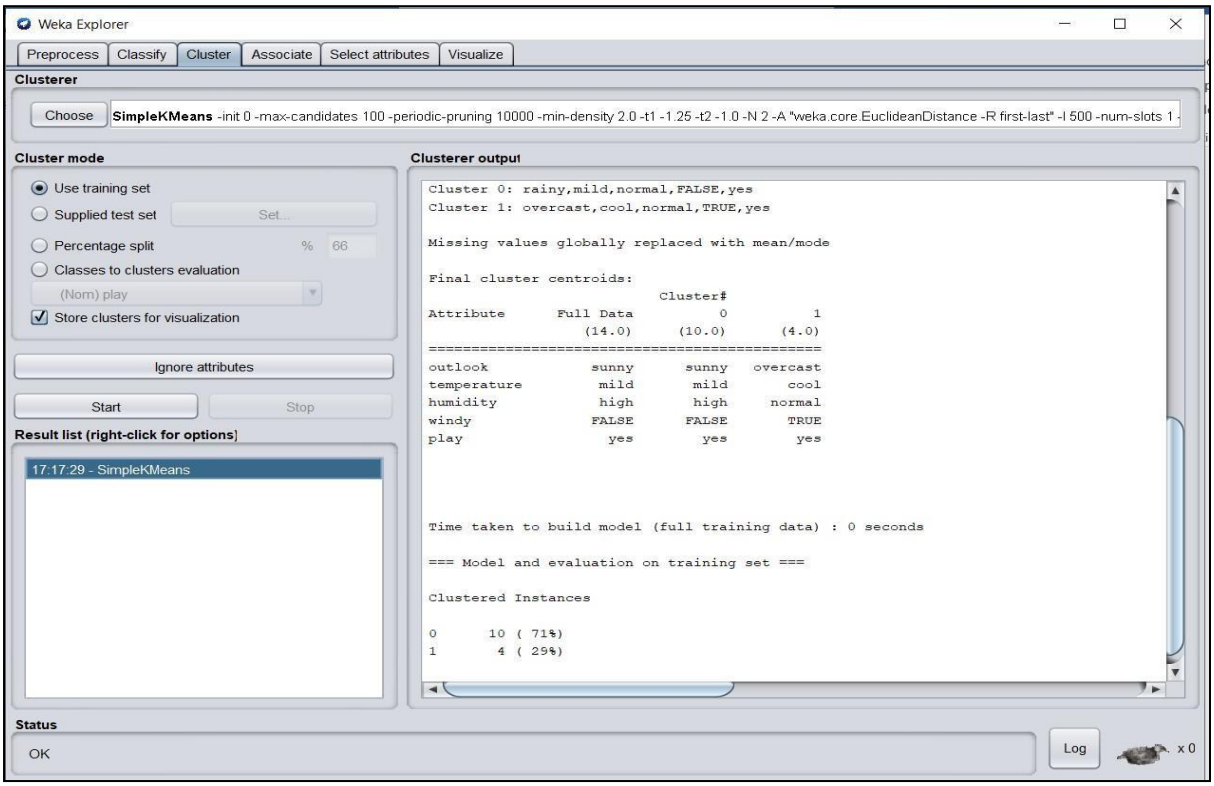
- **Explorer Model:**

Passing Dataset:



Output of K Means Clustering in Explorer Mode:




Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
Choose
SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -
Cluster mode
Use training set
Supplied test set
Set...
Percentage split
%
66
Classes to clusters evaluation
(Nom) play
Store clusters for visualization
Ignore attributes
Start
Stop
Result list (right-click for options)
17:17:29 - SimpleKMeans
Clusterer output
Cluster 0: rainy,mild,normal,FALSE,yes
Cluster 1: overcast,cool,normal,TRUE,yes
Missing values globally replaced with mean/mode
Final cluster centroids:
Cluster#
Attribute Full Data 0 1
(14.0) (10.0) (4.0)
outlook sunny sunny overcast
temperature mild mild cool
humidity high high normal
windy FALSE FALSE TRUE
play yes yes yes
Time taken to build model (full training data) : 0 seconds
=== Model and evaluation on training set ===
Clustered Instances
0 10 (71%)
1 4 (29%)
Status
OK
Log
x 0