Sanjivani Rural Education Society's
Sanjivani College of Engineering, Kopargaon-423 603
(An Autonomous Institute, Affiliated to Savitribai Phule Pune University, Pune)
NACC 'A' Grade Accredited, ISO 9001:2015 Certified

Department of Computer Engineering
(NBA Accredited)

Subject- Laboratory Practice II(410247)
Data Mining & Warehousing LAB(410247)
Lab Assignment 2- Visualize the clusters using suitable tool (WEKA) .

Prof. T.Bhaskar

Assistant Professor

Google-Site: https://sites.google.com/view/bhaskart/ug-notes/datamining-warehousing

Moodle-Site: https://proftbhaskar.gnomio.com/course/view.php?id=3 **(Log in as Guest)**

DMW YouTube Playlist: https://tinyurl.com/DMW-Bhaskar

Assignment Problem Statement

Consider a suitable dataset. For clustering of data instances in different groups, apply different clustering techniques (minimum 2). Visualize the clusters using suitable tool.

Relevant Theory

CLUSTERING: - Clustering is a task of assigning a set of objects into groups called as clusters. Clustering is also referred as cluster analysis where the objects in the same cluster are more similar to each other than to those objects in other clusters.

Clustering is the main task of Explorative Data mining and is a common technique for statistical data analysis used in many fields like machine learning, pattern recognition, image analysis, bio informatics etc…

Cluster analysis is not an algorithm but is a general task to be solved.

Clustering is of different types like hierarchical clustering which creates a hierarchy of clusters, partial clustering, and spectral clustering.

SimpleK-Means: -

It is a method of cluster analysis called as partial cluster analysis or partial clustering.

K-Means clustering partition or divides **n** observations into K clusters.

Each observation belongs to the cluster with the nearest mean.

K-means clustering is an algorithm to group the objects based on attributes/features into K number of groups where K is positive integer.

K-Means clustering is used in different types of applications like pattern recognition, artificial intelligent, image processing, etc…

Now Open the **WEKA GUI** Chooser from start menu

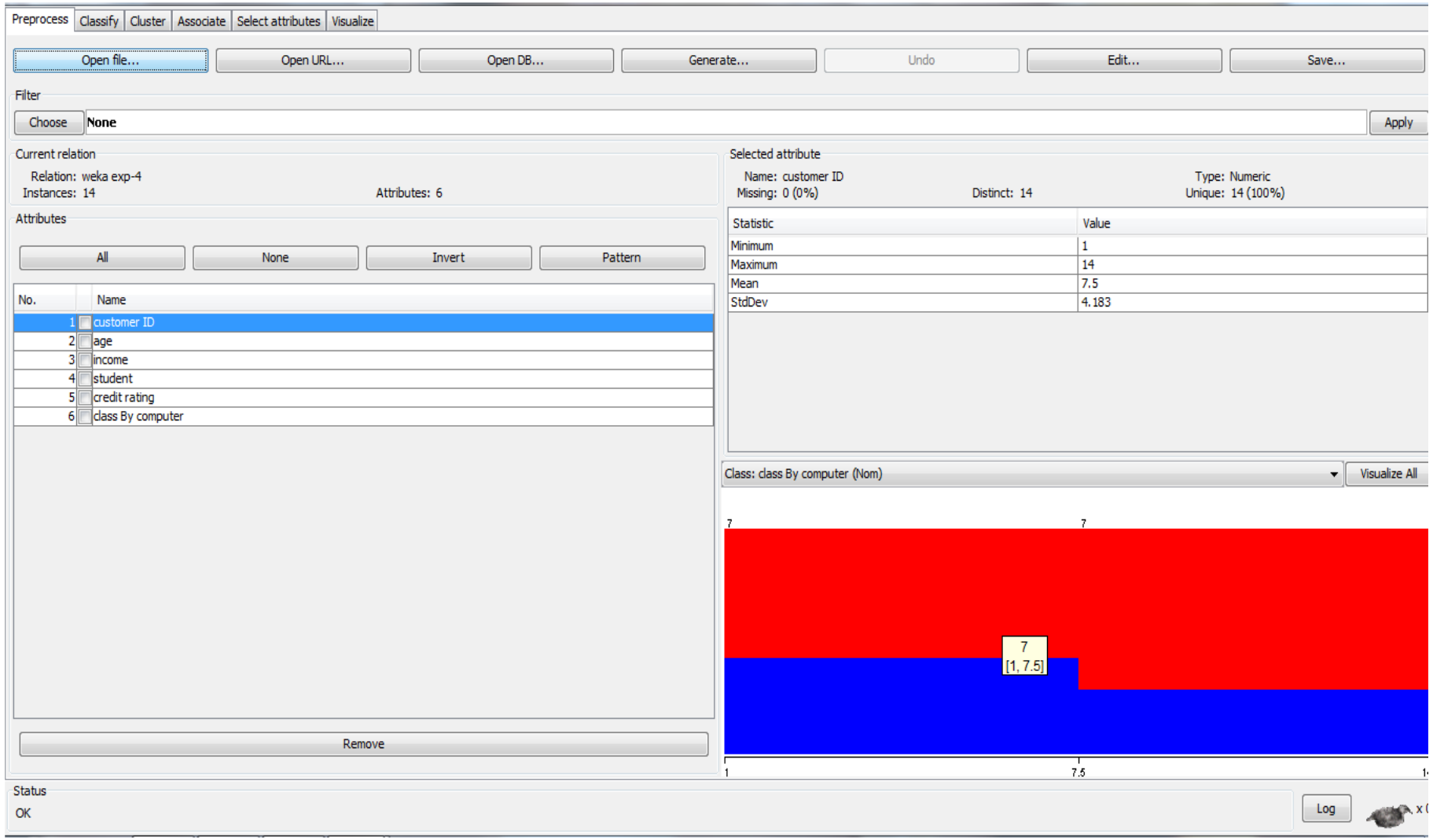
all programs and click on the EXPLORER button.

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

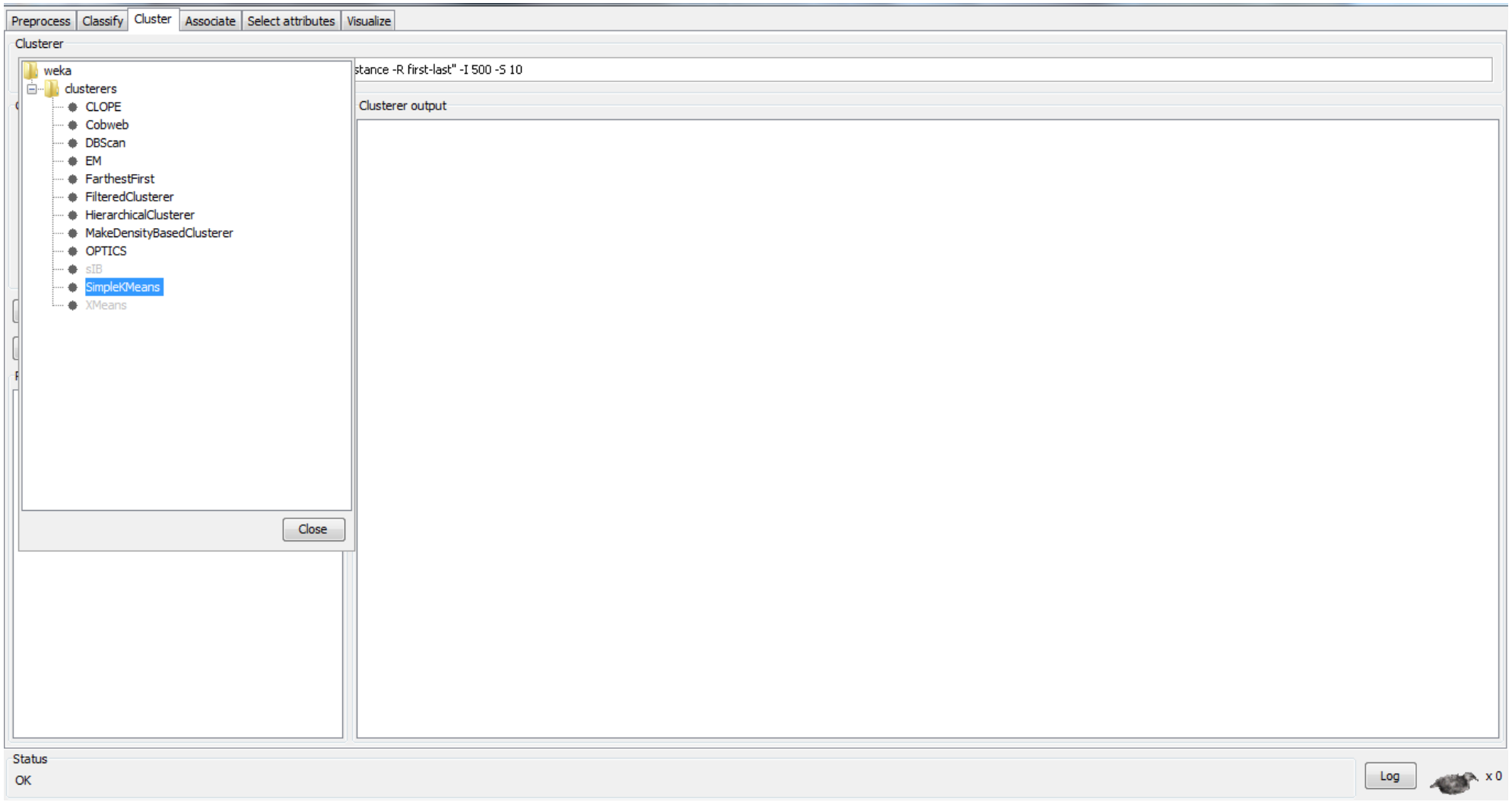
CLUSTER.CSV

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

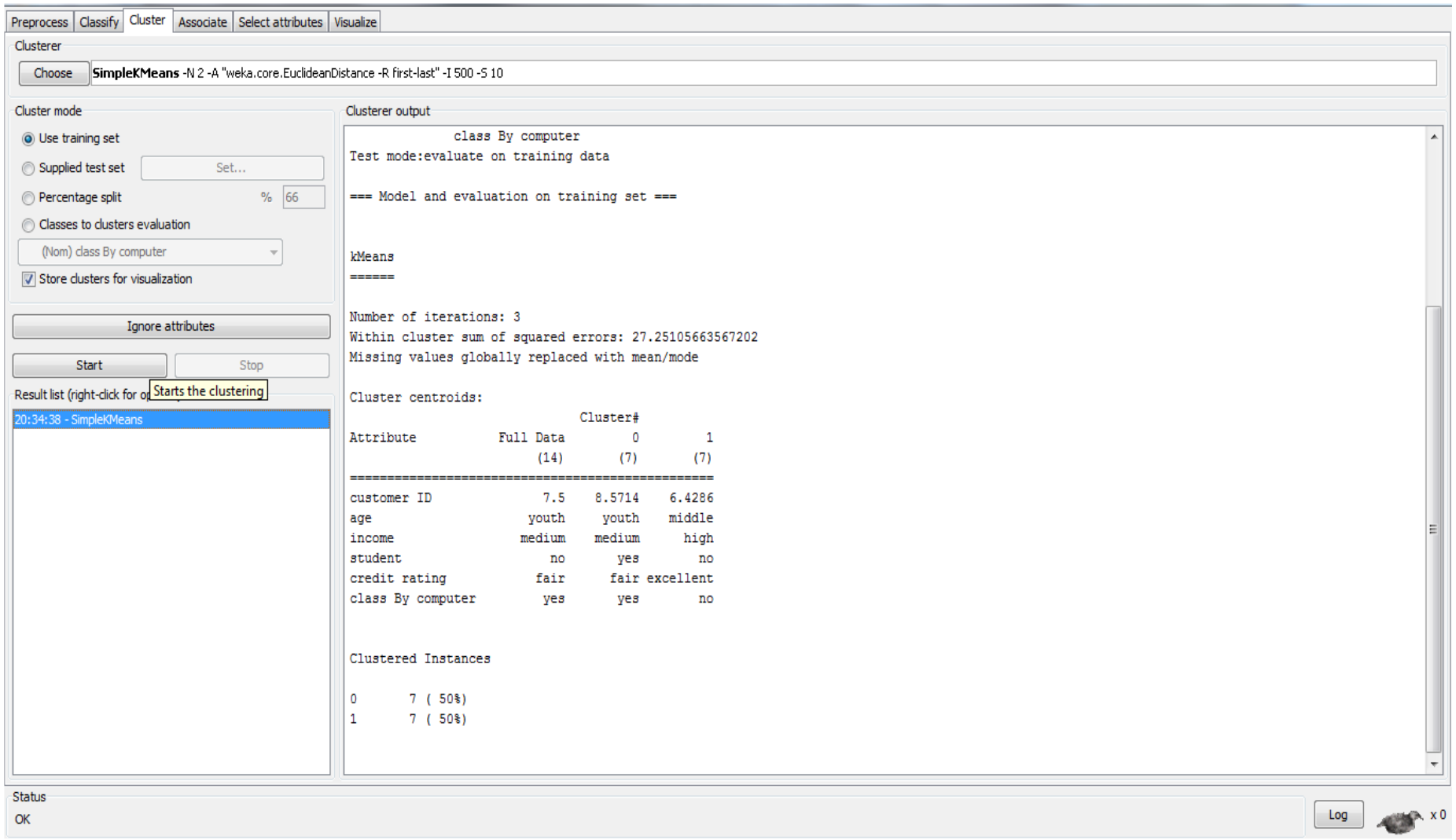
Load The CLUSTER.CSV file in Weka Tool



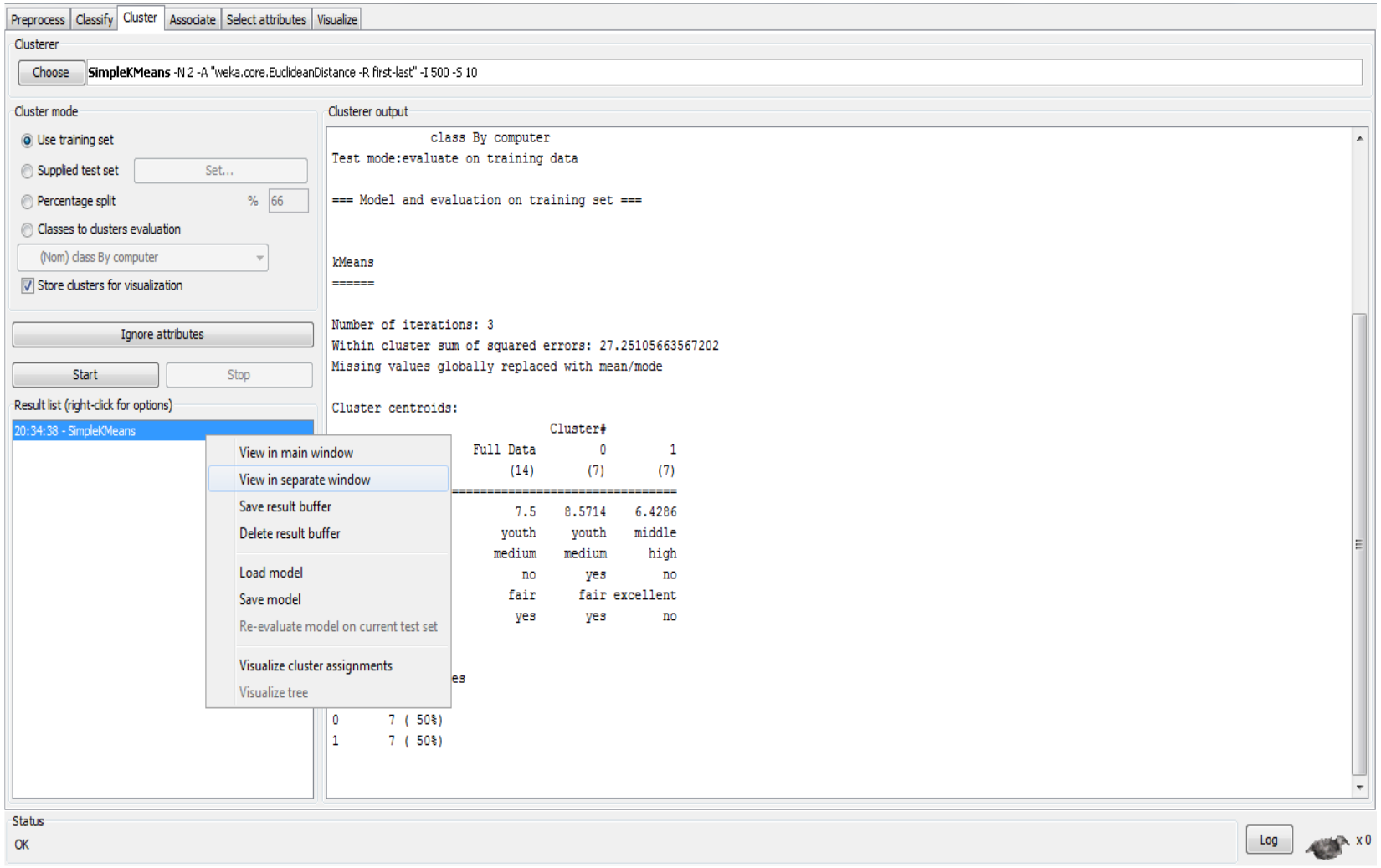
Selecting Simple Kmeans



Selecting Use Training Set



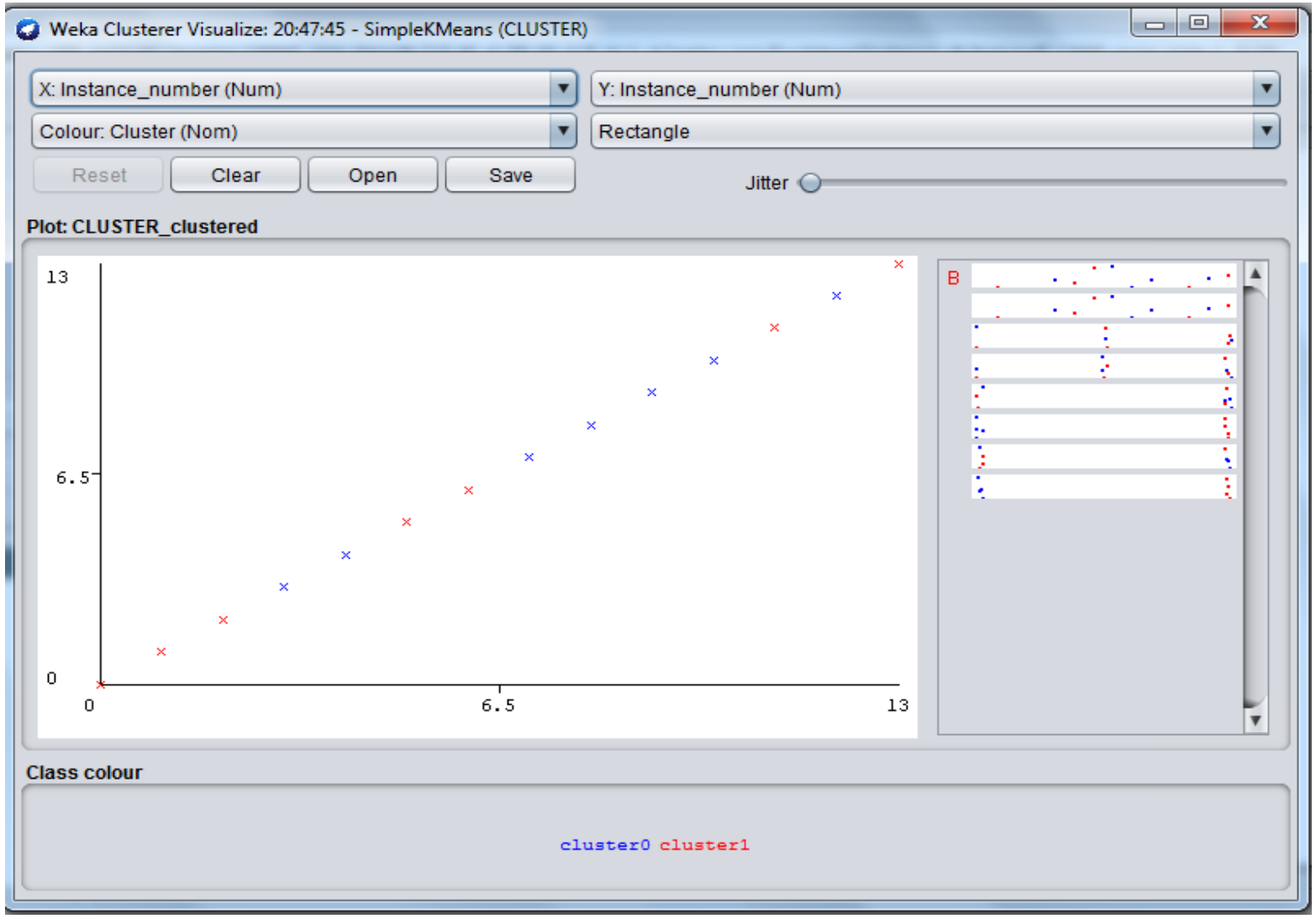
SELECTING THE OPTION "VIEW IN SEPARATE WINDOW"



The output is viewed in a separate window is as follows:

```
=== Run information ===
Scheme:weka.clusterers.SimpleKMeans -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -S 10
Relation: weka exp-4
Instances:14
Attributes:6
customer ID
age
income
student
credit rating
class By computer
Test mode:evaluate on training data
=== Model and evaluation on training set ===
kMeans
======
Number of iterations: 3
Within cluster sum of squared errors: 27.25105663567202
Missing values globally replaced with mean/mode
Cluster centroids:
Cluster#
Attribute Full Data 0 1
(14) (7) (7)
==============================================
```

Visualize the Cluster



Cluster data using the Farthest First algorithm.

=== Run information ===

Scheme: weka.clusterers.FarthestFirst -N 2 -S 1

Relation: CLUSTER

Instances: 14

Attributes: 6

customer ID

age

income

student

credit rating

class By computer

Test mode: evaluate on training data

=== Clustering model (full training set) ===

FarthestFirst

==========

Cluster centroids:

Cluster 0

12.0 middle medium no excellent yes

Cluster 1

1.0 youth high no fair no

Time taken to build model (full training data) : 0 seconds

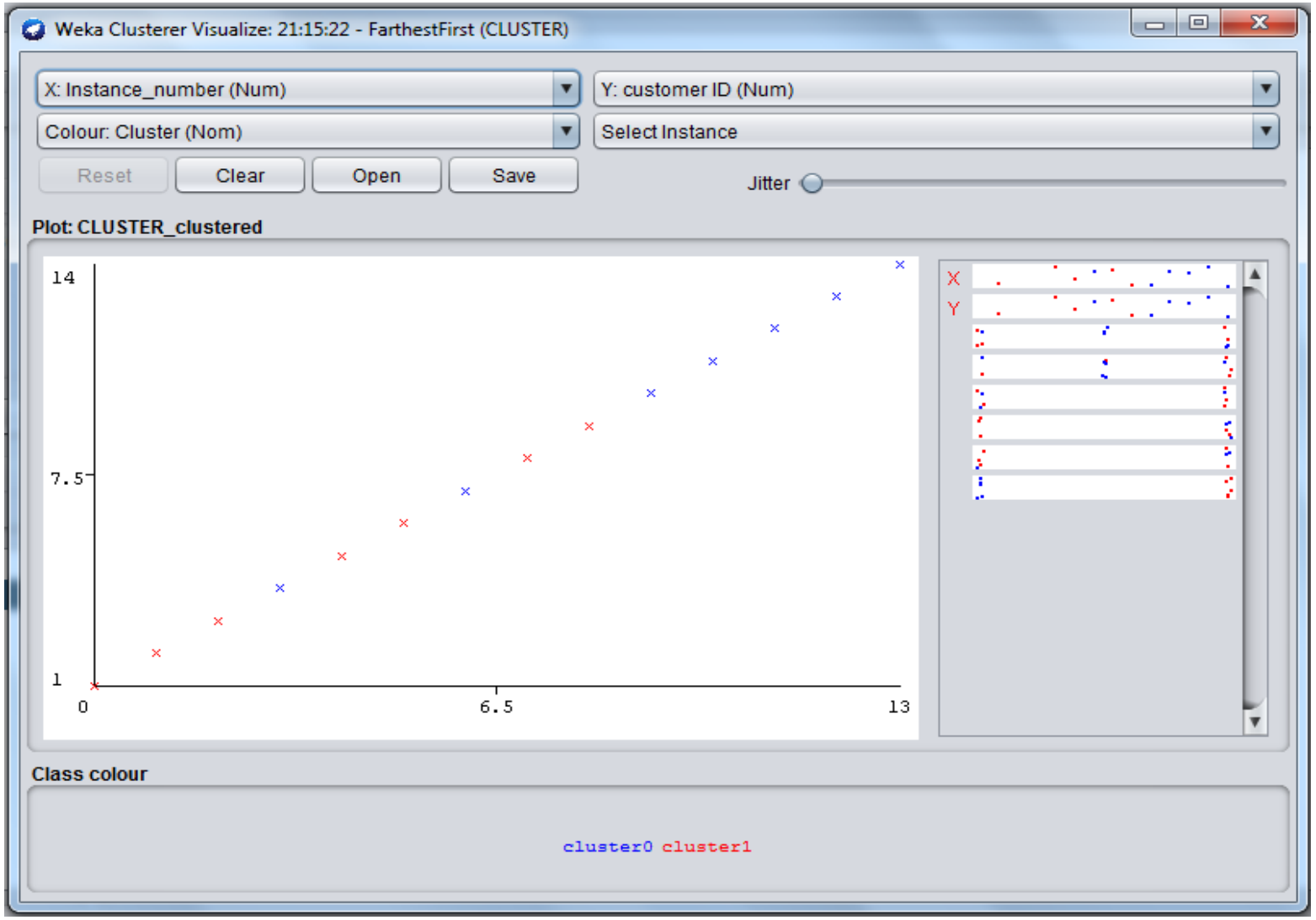
=== Model and evaluation on training set ===

Clustered Instances

0 7 (50%)

1 7 (50%)

Visualize the Cluster



Conclusion:

Created the clusters visualization.

Suggested Readings

Text Books:

Sr. No.	Title of Book	Authors	Publication House
1	Data Mining: Concepts and Techniques	Han, Jiawei Kamber, Micheline Pei and Jian	Elsevier Publishers
2	Reinforcement and Systemic Machine Learning for Decision Making	Parag Kulkarni	Wiley-IEEE Press

Reference Books:

Sr. No.	Title of Book	Authors	Publication House
1	Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More	Matthew A. Russell	Shroff Publishers
2	Social Network Analysis for Startups:Finding connections on the social web	Maksim Tsvetovat, Alexander Kouznetsov	Shroff Publishers

For further queries & doubts:
bhaskarcomp@sanjivani.org.in

The Material Used in this Presentation has been compiled from various Sources: Book by Data Mining: Concepts and Techniques by Han, Jiawei Kamber, Micheline Pei and Jian Elsevier Publishers & Other Books ,Lecture Notes, Tutorials & Online Resources.