Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC Accredited with "A" Grade (CGPA : 3.18)

A.Y. 2022-2023

Name : Siddharth Unny SAP-ID : 60004200080 Batch : A4

LAB EXPERIMENT NO. 01

Aim : Perform data Pre-processing task using Weka data mining tool

Theory :

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

Tasks performed through Weka:

Preprocessing

Classification

Clustering

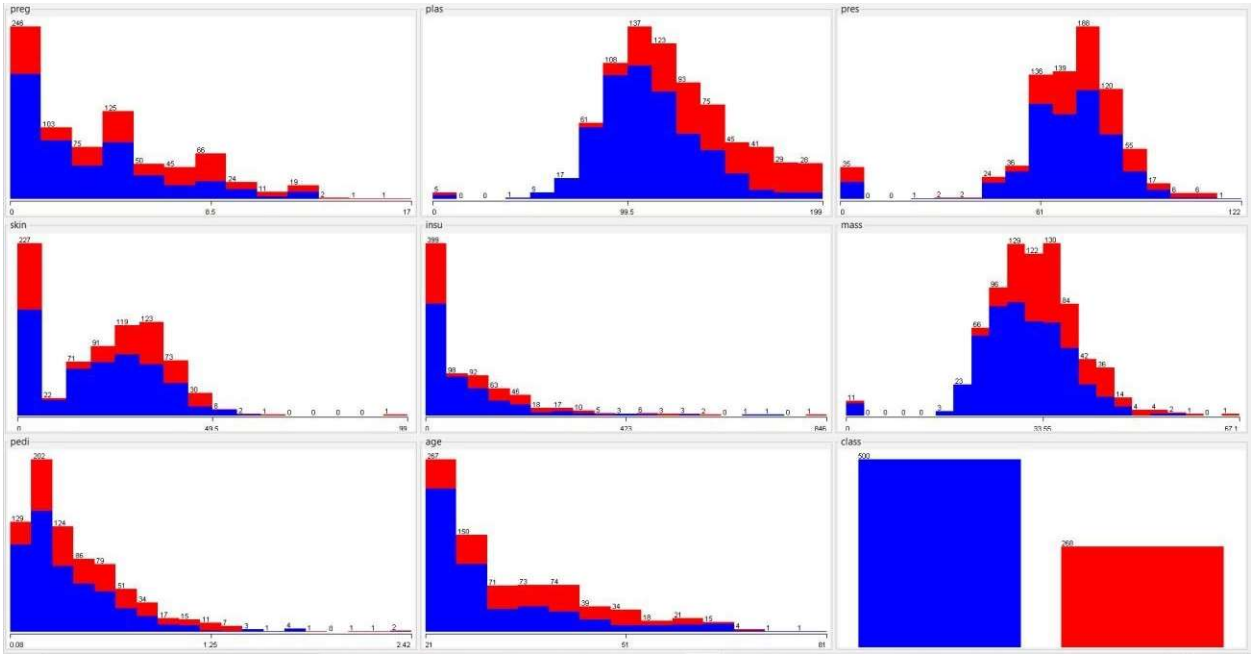
Association Rule

Select Attributes

Visualization

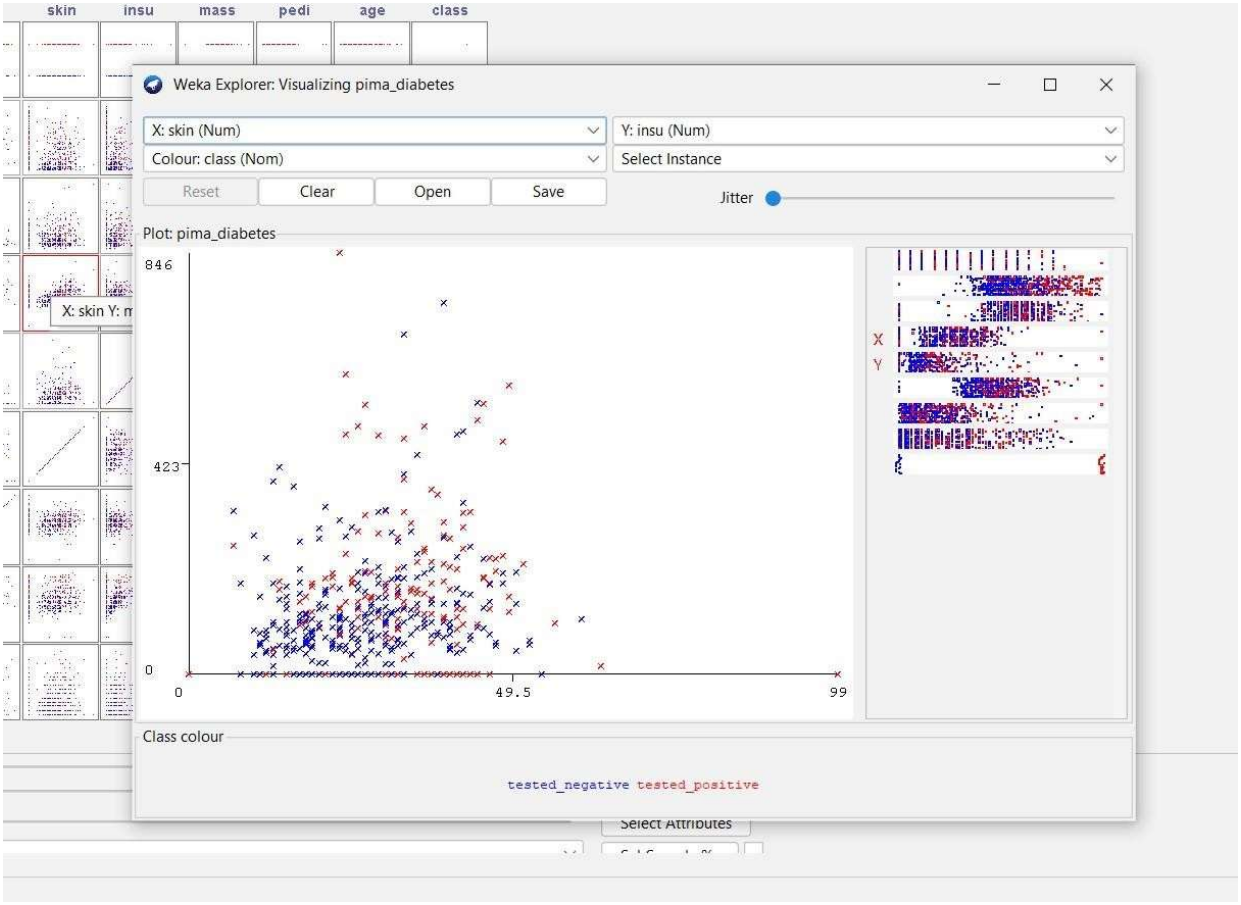
Output :

Visualization of all attributes

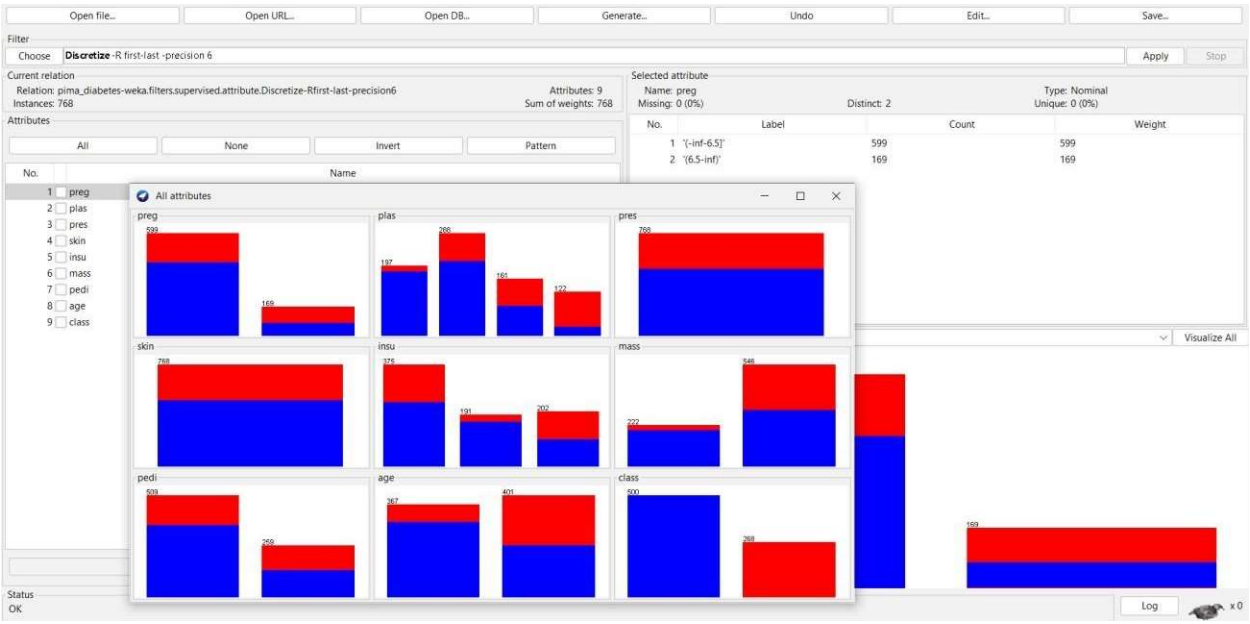


Scatter plot for all attributes

From the plot, we can see that attribute insu and skin are highly correlated along with minor correlation in plas, mass, pedi, age and class attributes.


skin
insu
mass
pedi
age
class
Weka Explorer: Visualizing pima_diabetes
X: skin (Num)
Y: insu (Num)
Colour: class (Nom)
Select Instance
Reset
Clear
Open
Save
Jitter
Plot: pima_diabetes
846
423
0
0
49.5
99
X
Y
Class colour
tested_negative tested_positive
Select Attributes

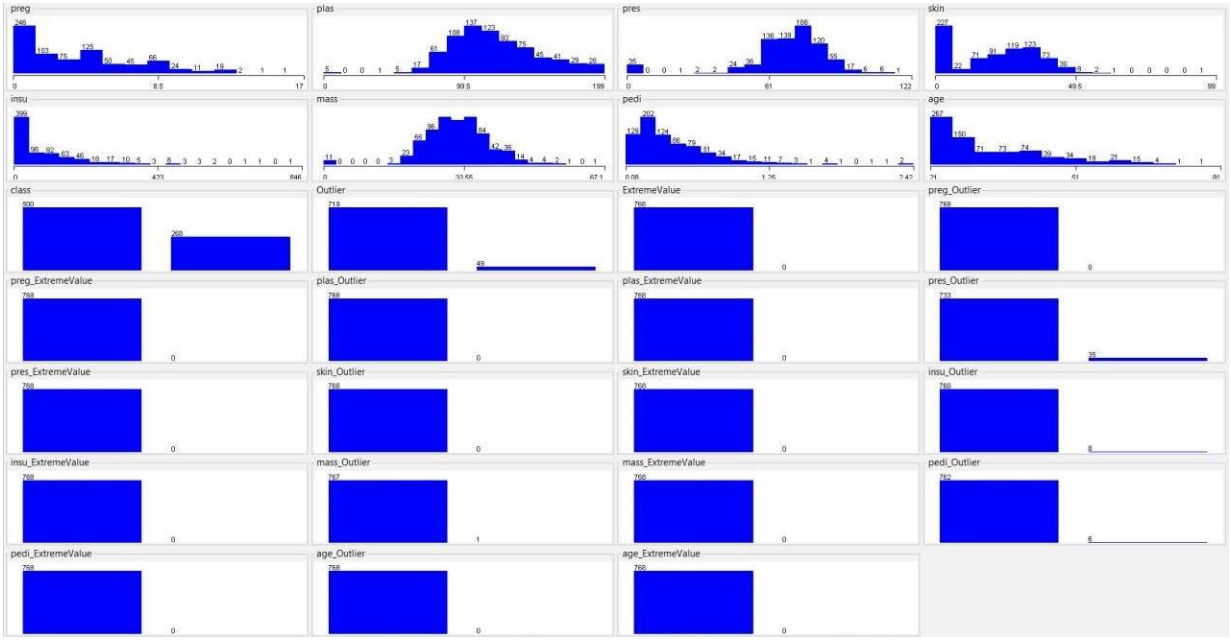
Supervised Filter – Discretization



Unsupervised Filter – Discretization

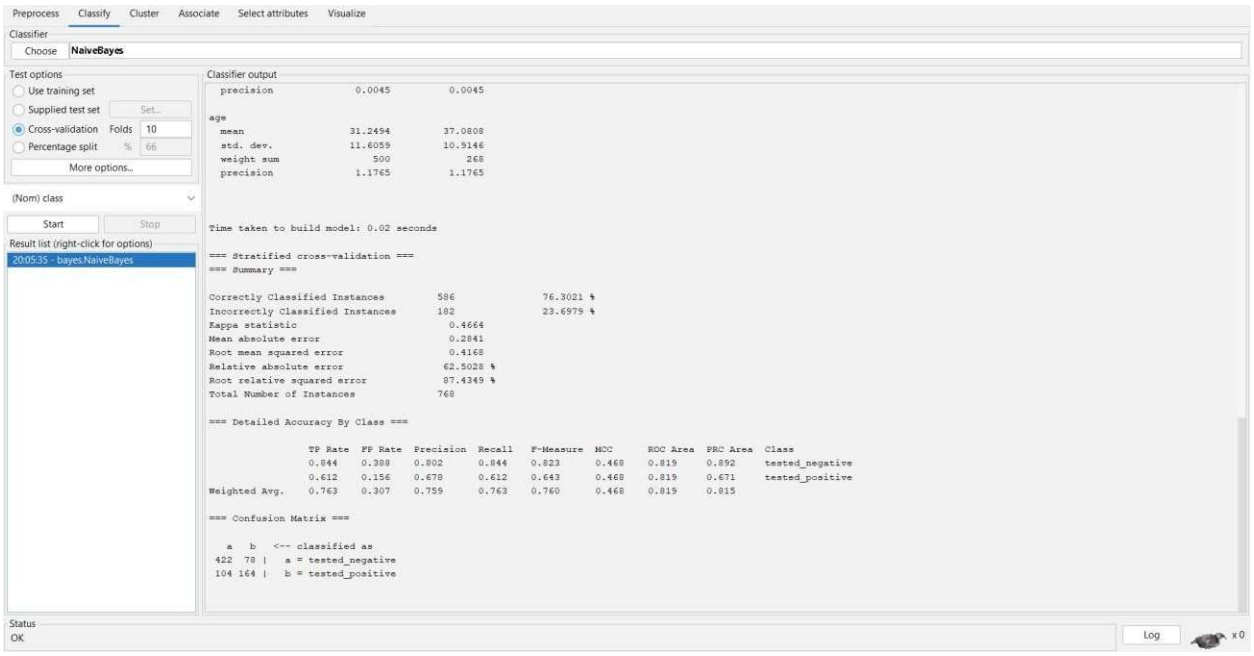
Supervised Filter – AttributeSelection

Unsupervised Filter – InterQuartileRange

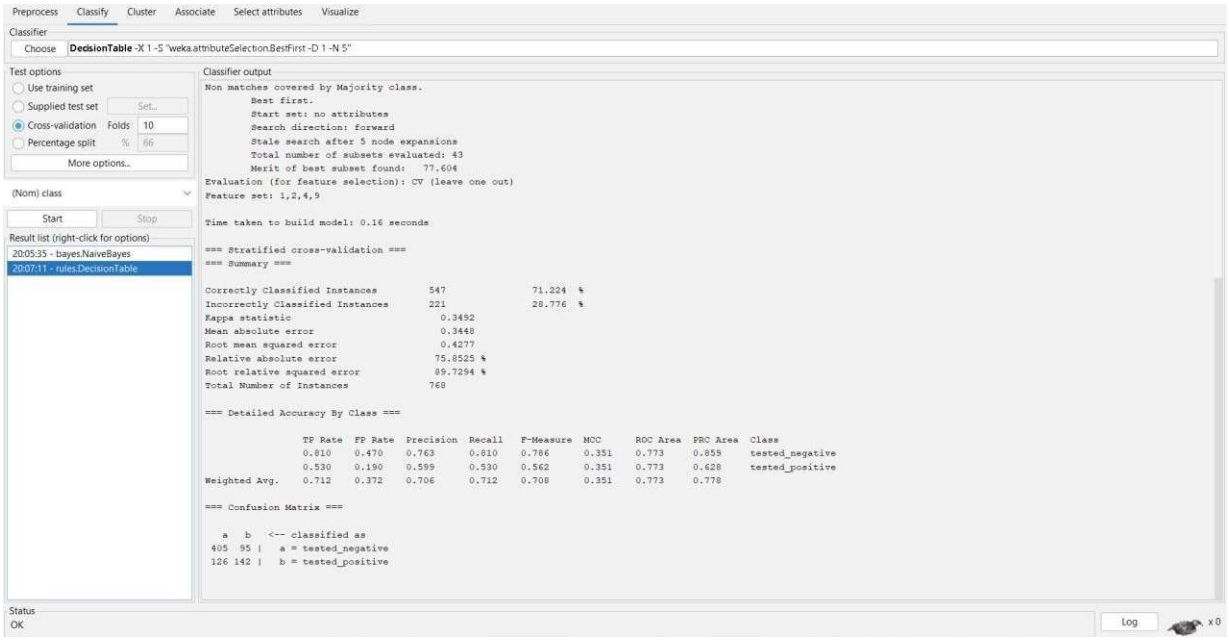


Classification :

Naïve Bayes

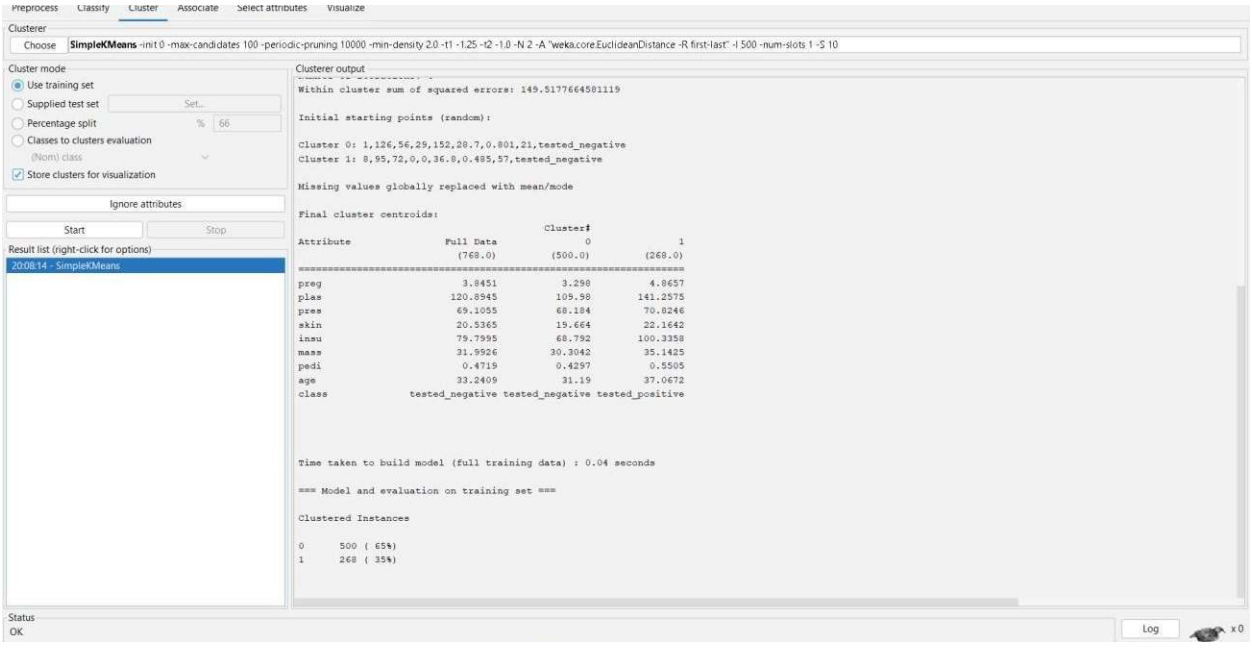


Decision Table:

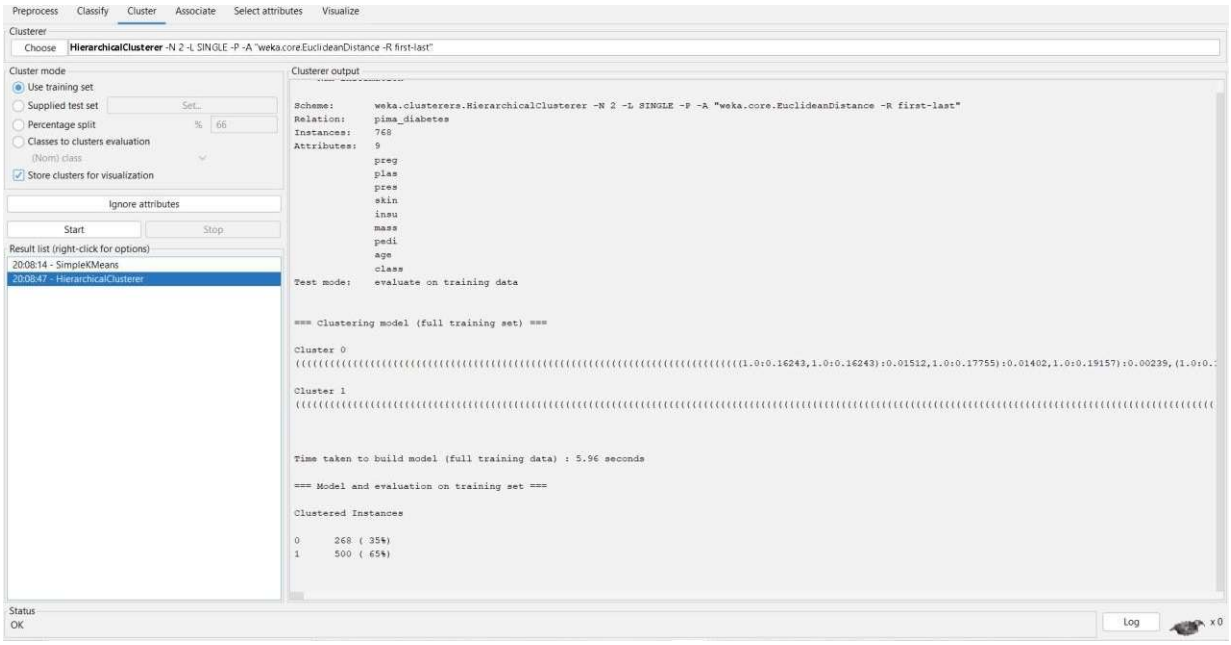


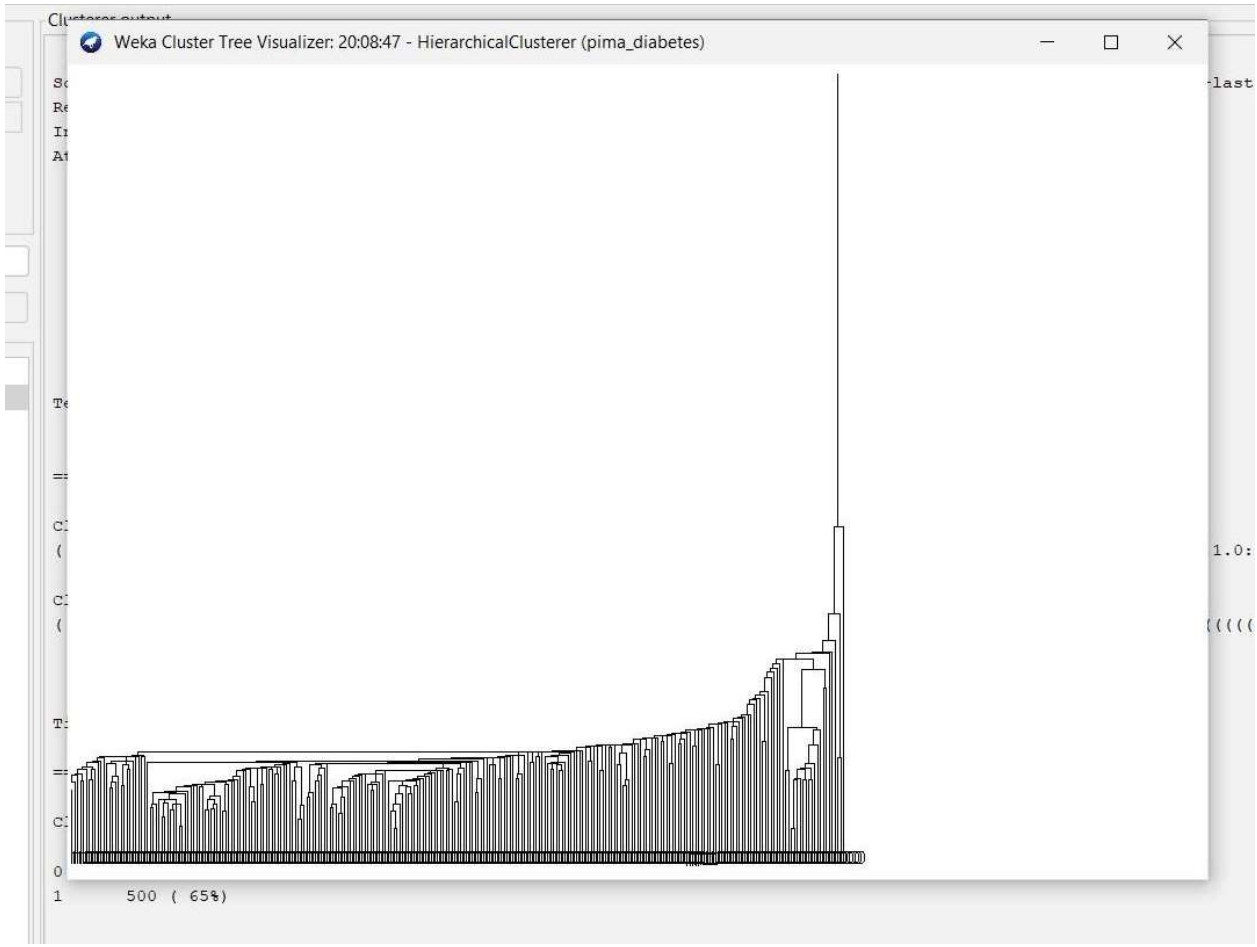
Clustering :

Simple K-Means



Hierarchical Clustering:

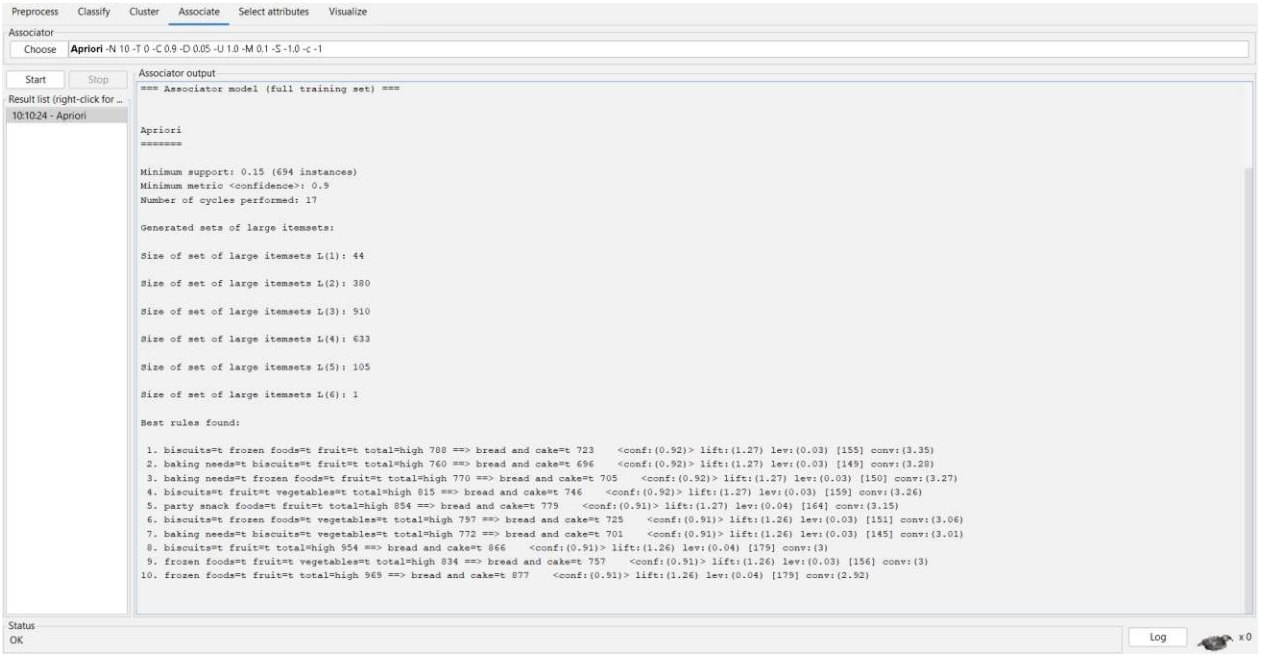




In clustering based on the attribute needed to be predicted random amount of clusters are formed, in this case there are two clusters 0 and 1 which denote the tested_negative and tested_positive values of attribute 'class' based on which samples from each cluster are passed in the model for training and the corresponding output displays the clusters with count of instances present in each cluster.

Since there is no association between the attributes, associate Apriori can't be used here.

Dataset – Supermarket – Apriori



Here we can see that attributes which associate with attributes bread and cake allowing the owner to understand which items should be placed together and such.