**Lab 3 Data Mining and Clustering**

**Group 10**

**Clustering**

In this part we tried different clustering techniques which are K-means, Density Based k-means, EM Algorithm and the Hierarchical Clustering algorithm. In this case EM algorithm performed the best with misclassification rate of 42.7% which is good as compared to others but is not particularly good as it seems to be random guessing. The problem is that as the clustering algorithms are not designed to classify observations so the objective function is completely different from the one we would have by a classification algorithm. The Clustering algorithm tries to minimize the intra-cluster distances and maximize inter-cluster distances, also the data is categorical which is sensitive for distance measure and the algorithm will have problems trying to minimize distance.

**Association Analysis**

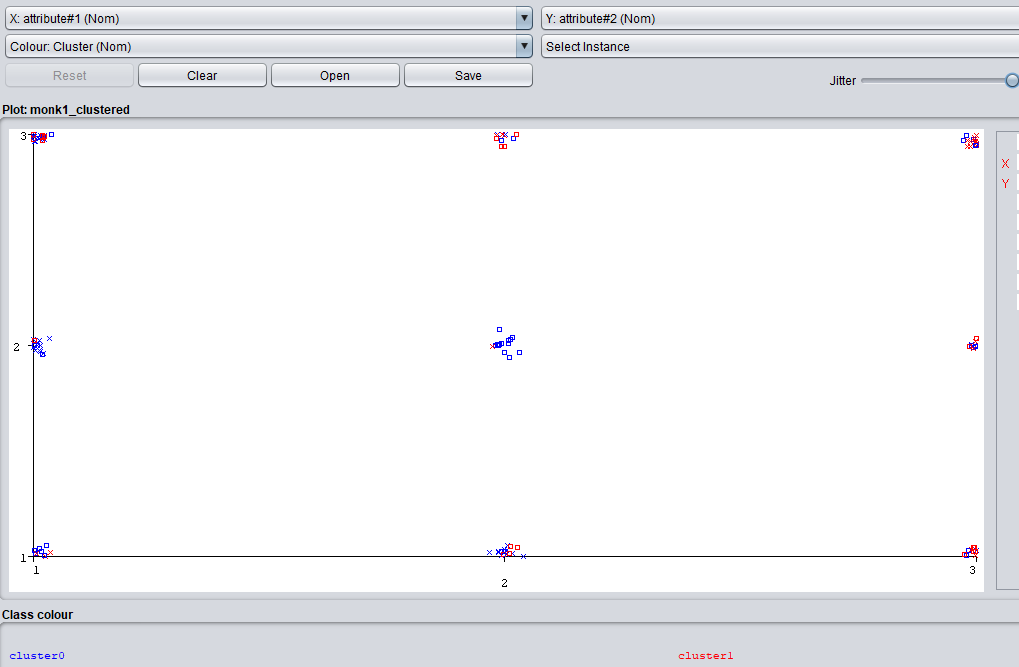
In this task I used Apriori Algorithm with use a minimum support of 0.05 and a maximum number of rules of 19. We selected following rules by analyzing the data set. Also we have to select rules that accurately predict class 1 i.e. an instance is assigned to class 0 if it is not assigned to class 1 as mentioned in the task. We by observing selected the following 4 rules from the 19 which are:

C:\Users\Fahad Hameed\Desktop\a.PNG

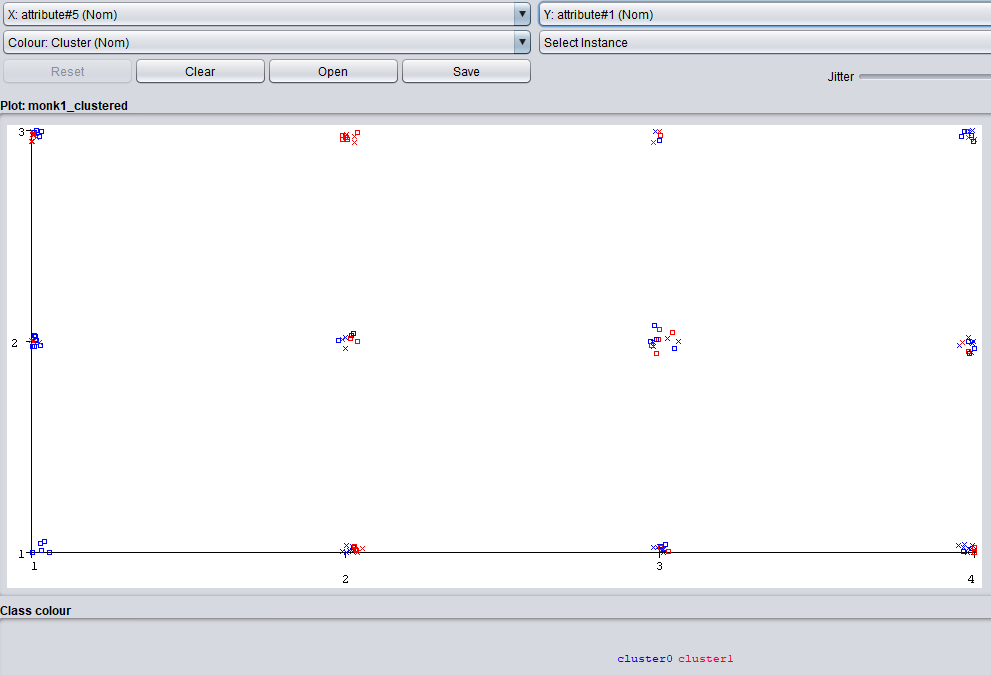
C:\Users\Fahad Hameed\Desktop\b.PNG

C:\Users\Fahad Hameed\Desktop\c.PNG

C:\Users\Fahad Hameed\Desktop\d.PNG



**Attribute#1 on X axis, Attribute#2 on Y axis**



**Attribute#5 on X axis, Attribute#1 on Y axis**

We can see that using Association analysis we are able to find the clusters in the data and the combining them to get separable classes. Whereas the clustering algorithm cannot do it since they do not have ability for clustering and just minimize distance within cluster.