



## **Experiment - 3.1**

Student Name: Pankaj Singh Kanyal UID: 20BCS6668

Branch: AIML Section/Group: 20AML-4B

Semester: 5 Date:

Subject Name: Data mining & warehousing lab Subject Code: 20CSF-333\_20AML-4

#### 1. Aim/Overview of the Practical

Write a procedure for Clustering Customer data using Simple KMeans Algorithm.

#### 2. Task to be done

- 1) Create the weather table using notepad
- 2) Create data and use the evaluation and visualize tab to flow the data using knowledge flow.

### 3. Program Code:

#### **Procedure/ steps**

- 1) Make an .arff file containing customer data.
- 2) After that the file is saved with .arff file format.
- 3) Minimize the .arff file and then open Start à Programs à weka-3-4.
- 4) Click on weka-3-4, then the Weka dialog box is displayed on the screen.
- 5) In that dialog box there are four modes, click on explorer.
- 6) Explorer shows many options. In that click on 'open file' and select the .arff file
- 7) Click on the edit button which shows the buying table on weka.





#### 4. Arff Customer data file

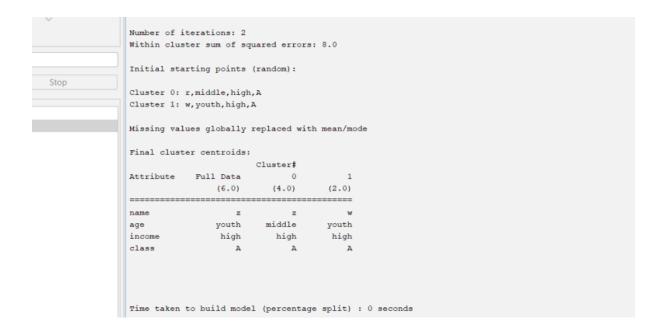
@relation customer
@attribute name {x,y,z,u,v,l,w,q,r,n}
@attribute age {youth,middle,senior}
@attribute income {high,medium,low}
@attribute class {A,B}
@data x,youth,high,A
y,youth,low,B
z,middle,high,A
u,middle,low,B
v,senior,high,A
l,senior,low,B
w,youth,high,A
q,youth,low,B
r,middle,high,A
n,senior,high,A

#### 4. Output

```
Clusterer output
Within cluster sum of squared errors: 14.000000000000004
Initial starting points (random):
Cluster 0: u,middle,low,B
Cluster 1: w, youth, high, A
Missing values globally replaced with mean/mode
Final cluster centroids:
                   Cluster#
Attribute Full Data (10.0)
                                  1
                        0
                             (6.0)
                      (4.0)
Time taken to build model (full training data) : 0 seconds
=== Model and evaluation on training set ===
Clustered Instances
      4 ( 40%)
      6 ( 60%)
```







#### 5. Observations

- We have created 2 clusters or groups using simple K Means.
- The dataset contains four attributes as name, age, income, class.
- There are total 10 records in the data set out of which 4 belong to cluster 0 and rest of 6 belong to cluster 1

#### 6. Result and Conclusion

Successfully implemented the Simple K-means in customer data.

## 7. Learning Outcomes

- 1. Learned to use knowledge flow in WEKA data mining software
- 2. Learned about Arff loaders and use them in knowledge flow
- 3. Learned and implemented cross validation techniques in weather dataset.

# Evaluation Grid (To be created as per the SOP and Assessment guidelines by the faculty):

	• /		
Sr.	Parameters	Marks Obtained	Maximum
No.			Marks
1.			
2.			
3.			

