

A  
Practical File  
On  
Data ware house and data mining

Paper Code: ITD08



Bachelor of Engineering  
(2016-2020)  
Computer Science Engineering

**Submitted By:**

Kavita Maurya

Batch-3

Group-1

2016UCO1579

# 1.

## AIM:

Write a procedure for Clustering Weather data using EM Algorithm.

## DESCRIPTION:

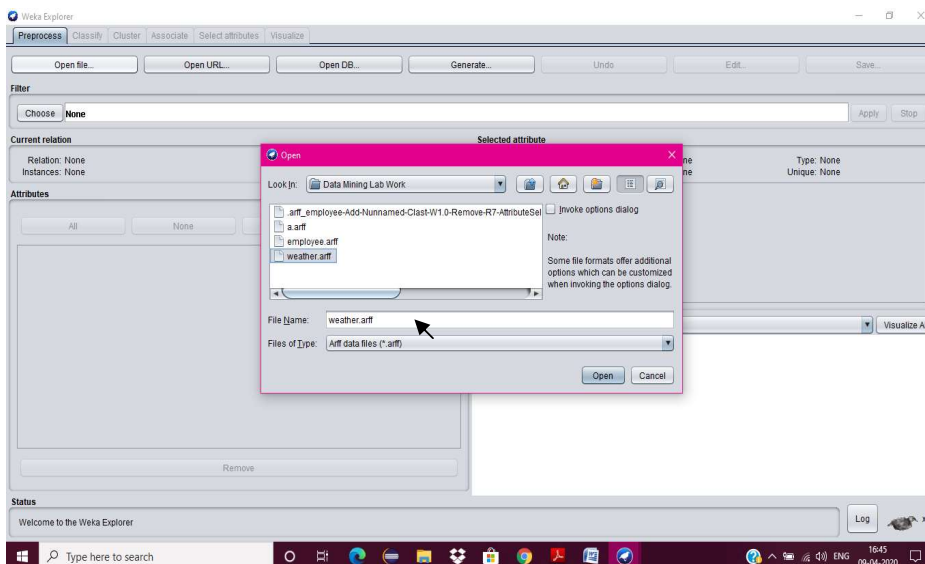
Cluster analysis or clustering is the task of assigning a set of objects into groups (called clusters) so that the objects in the same cluster are more similar (in some sense or another) to each other than to those in other clusters. Clustering is a main task of explorative data mining, and a common technique for statistical data analysis used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics.

## PROCEDURE:

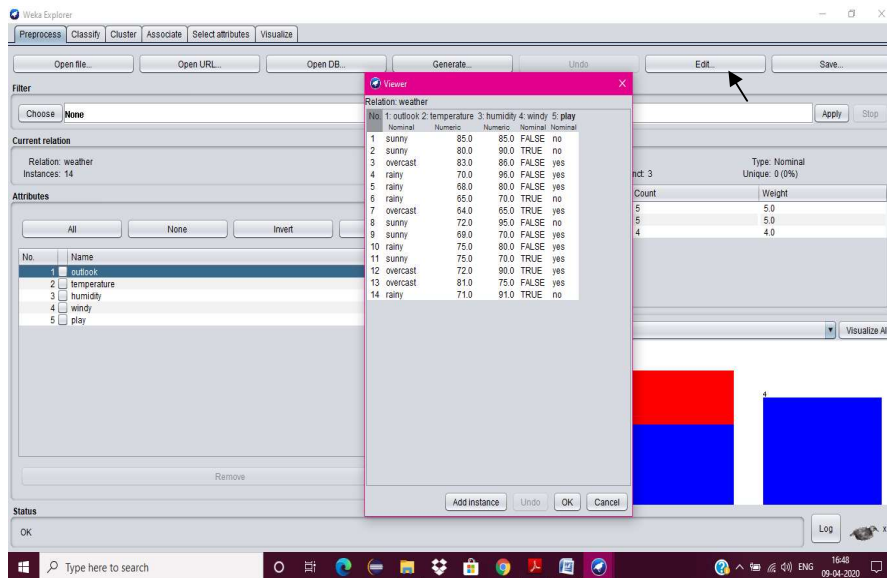
1. Create or download **weather.arff** file having weather data.
2. Click on weka and then click on explorer.



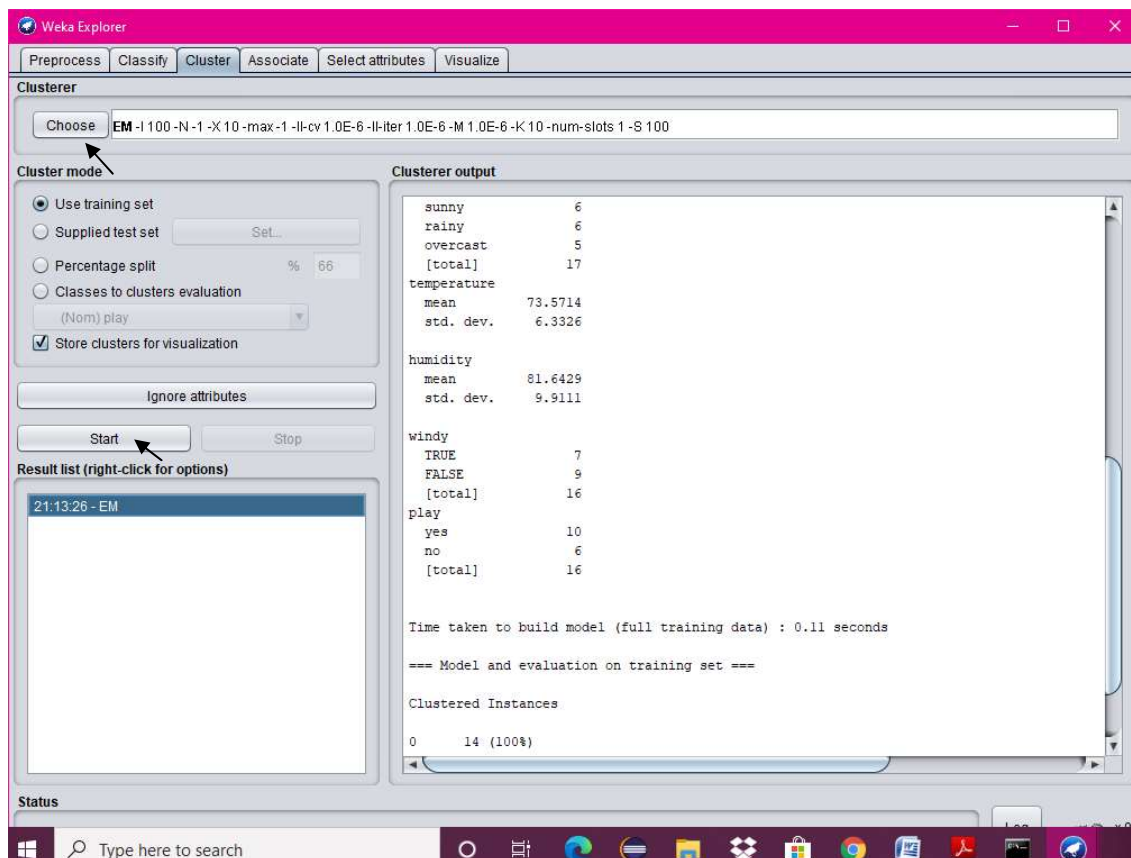
3. Explorer shows many options. In that click on 'open file' and select the arff file.



- Click on edit button which shows weather table on weka.



- Close the file.
- Click on Cluster menu. In this there are different algorithms are there.
- Click on Choose button and then select EM algorithm.
- Click on Start button and then output will be displayed on the screen.



## 2.

### AIM:

Write a procedure for Clustering Banking data using the farthest first Algorithm.

### DESCRIPTION:

Cluster analysis or clustering is the task of assigning a set of objects into groups (called clusters) so that the objects in the same cluster are more similar (in some sense or another) to each other than to those in other clusters. Clustering is a main task of explorative data mining, and a common technique for statistical data analysis used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics.

### PROCEDURE:

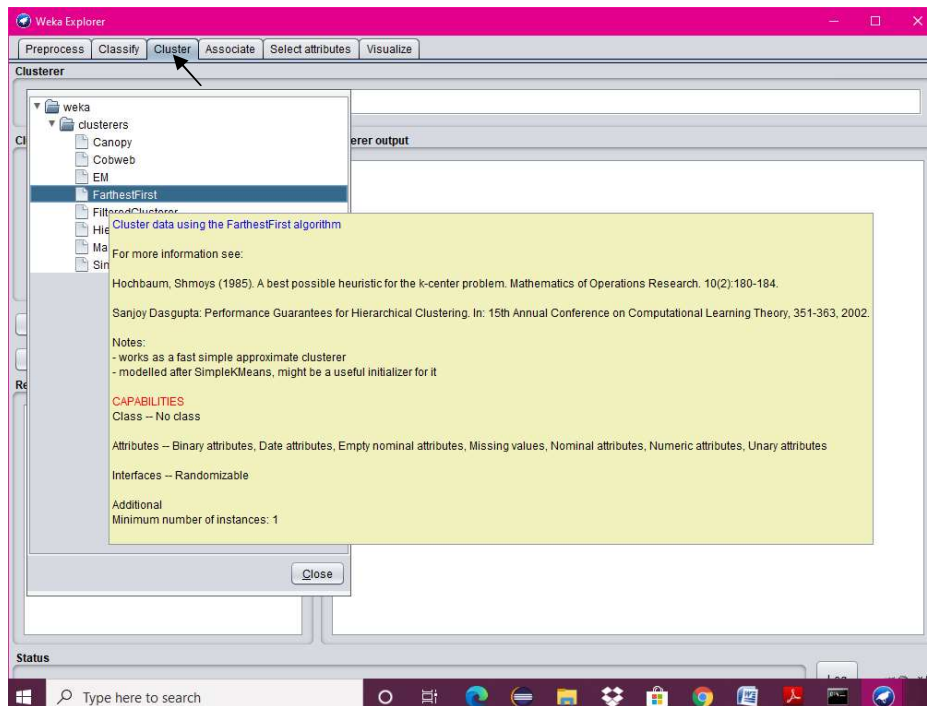
1. Create or download **banking.arff** file having banking data.
2. Click on weka and then click on explorer.



3. Explorer shows many options. In that click on 'open file' and select the arff file.
4. Click on edit button which shows banking table on weka.

No.	1: cust	2: accno	3: bankname	4: location	5: deposit
	Nominal	Nominal	Nominal	Nominal	Nominal
1	male	0101	sbi	hyd	yes
2	fem...	0102	hdfc	jmd	no
3	male	0103	sbh	antp	yes
4	male	0104	ab	pdtr	yes
5	fem...	0105	sbi	jmd	no
6	male	0106	ab	hyd	yes
7	fem...	0107	rbi	jmd	yes
8	fem...	0108	hdfc	kdp	no
9	male	0109	sbh	kdp	yes
10	male	0110	ab	jmd	no
11	fem...	0111	rbi	kdp	yes
12	male	0112	sbi	jmd	yes
13	fem...	0113	rbi	antp	no
14	male	0114	hdfc	pdtr	yes
15	fem...	0115	sbh	pdtr	no

5. Close the file.
6. Click on Cluster menu. In this there are different algorithms are there.
7. Click on Choose button and then select **FarthestFirst** algorithm.



8. Click on Start button and then output will be displayed on the screen.

