Girija.B

192311044

CSA1618 DWDM-DE

EXPERIMENT-23

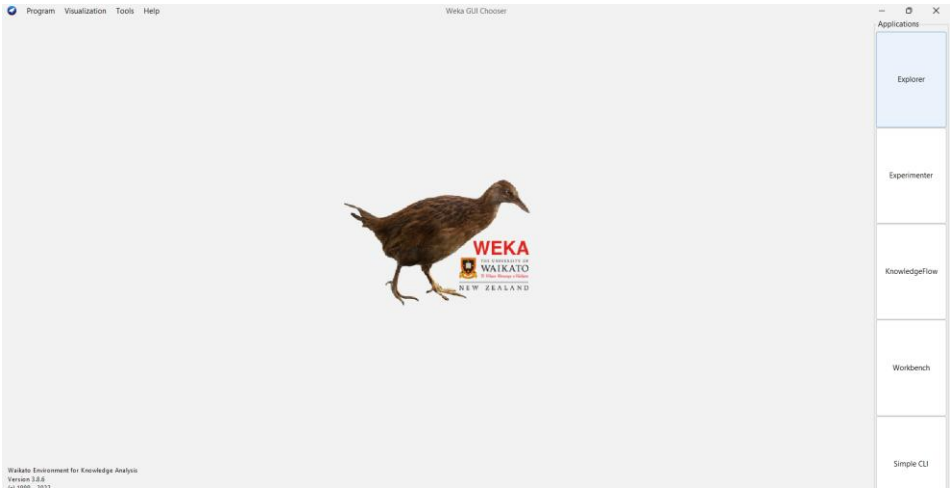
DATA SEGMENTATION BY EXPECTATION MAXIMISATION ALGORITHM THROUGH WEKA

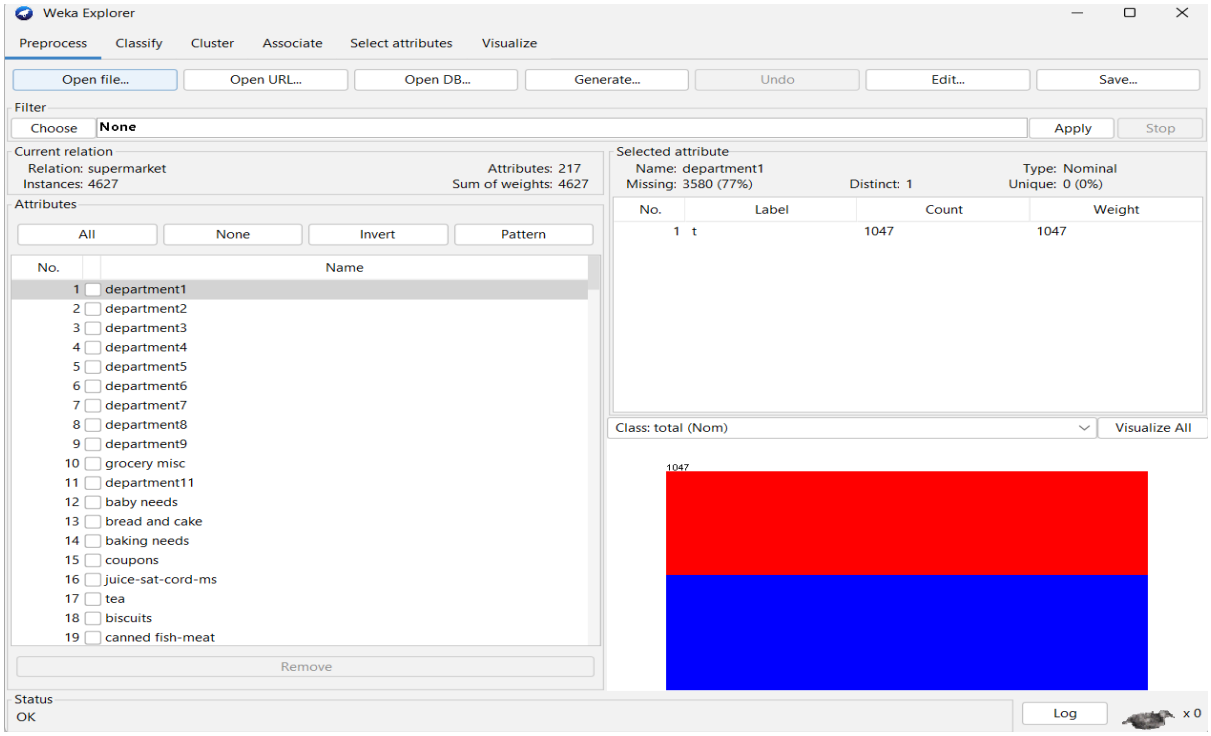
AIM:

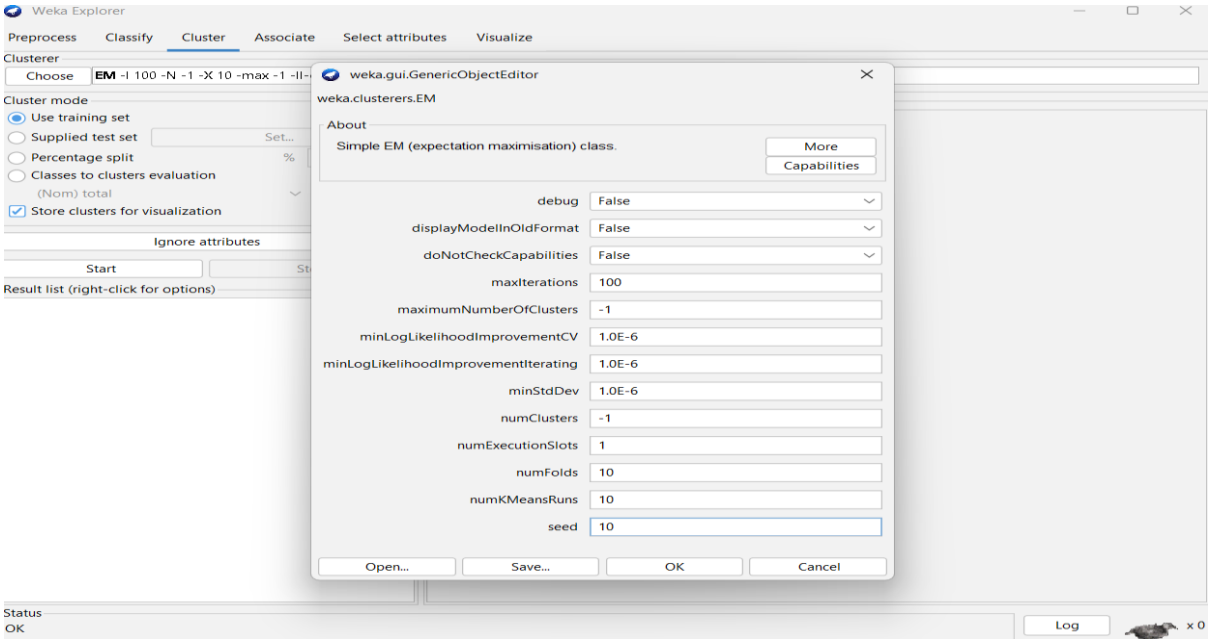
To create data segmentation by Expectation Maximisation algorithm through weka.

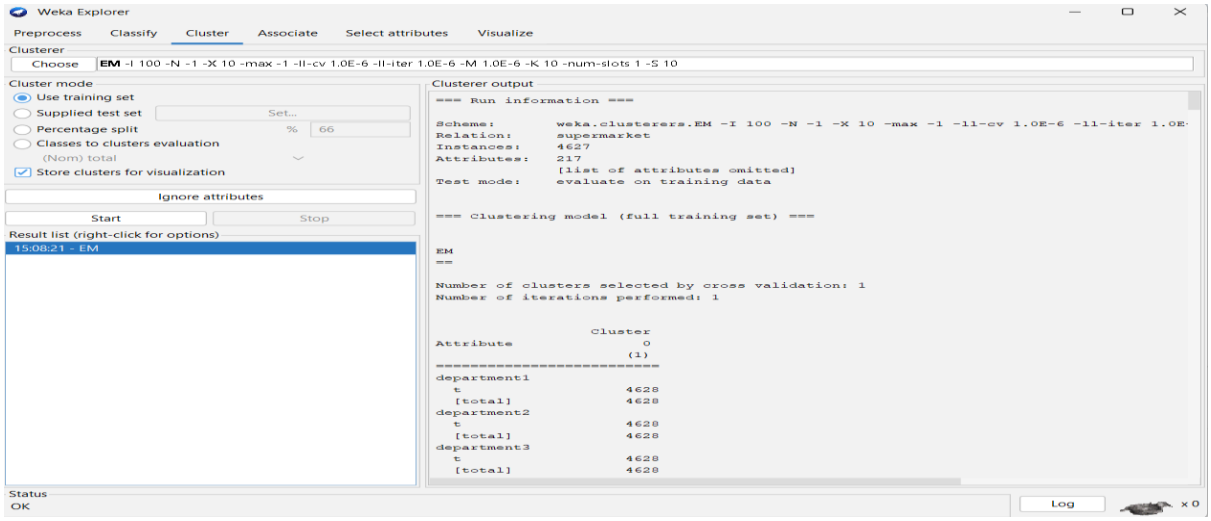
PROCEDURE:

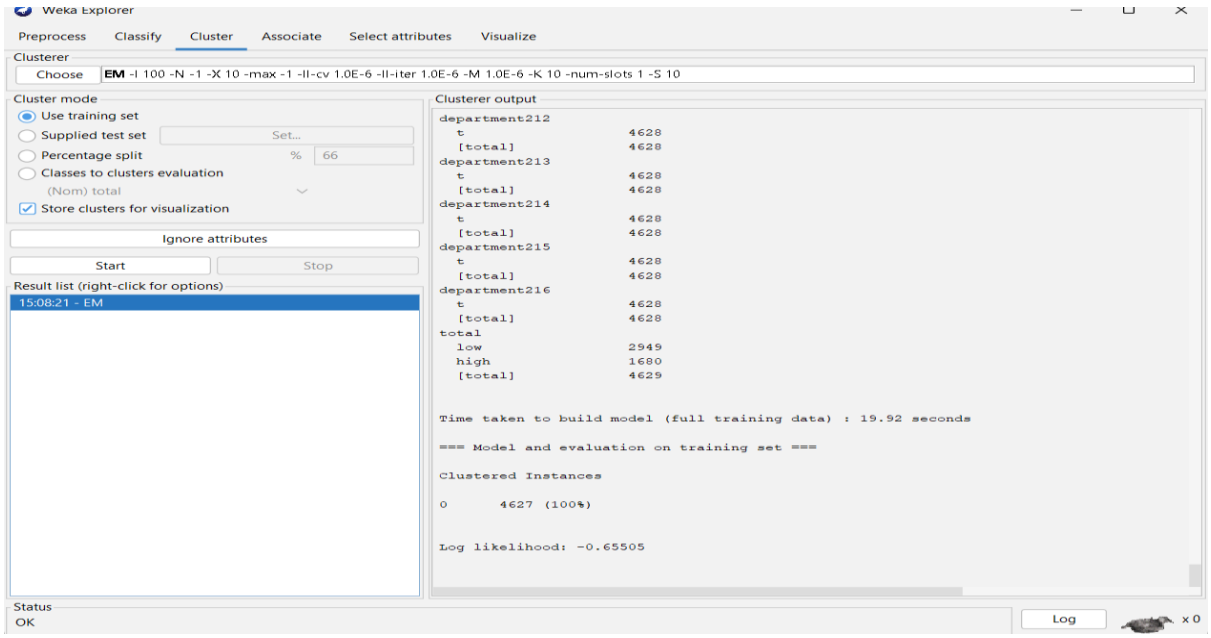
1. Download and install WEKA.
2. Open WEKA and Choose "Explorer" from the main menu.
3. Under Preprocess, Click on the open file button and select the dataset.
4. Click on the "Cluster" tab. In the Cluster mode section, select "Use training set".
5. Click "Choose" (next to the cluster algorithm) and Select EM (under weka. clusters).
6. Click on **"EM"** to configure it: **numClusters (-1 for automatic selection)** → WEKA will automatically determine the optimal number of clusters. **MaxIterations** → Set to 100 (default) or increase for better accuracy. **Seed** → Keep a fixed value (e.g., 10) for reproducibility.
7. Click **"OK"** and then **"Start"** to run the EM clustering. WEKA will display cluster assignments and statistics.
8. Click "Visualize" to see how the clusters are distributed. Save the file.




Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
None
Apply
Stop
Current relation
Relation: supermarket
Instances: 4627
Attributes: 217
Sum of weights: 4627
Attributes
All
None
Invert
Pattern
No.
Name
1 department1
2 department2
3 department3
4 department4
5 department5
6 department6
7 department7
8 department8
9 department9
10 grocery misc
11 department11
12 baby needs
13 bread and cake
14 baking needs
15 coupons
16 juice-sat-cord-ms
17 tea
18 biscuits
19 canned fish-meat
Remove
Selected attribute
Name: department1
Missing: 3580 (77%)
Distinct: 1
Type: Nominal
Unique: 0 (0%)
No.
Label
Count
Weight
1 t
1047
1047
Class: total (Nom)
Visualize All
1047
Status
OK
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
Choose
EM -I 100 -N -1 -X 10 -max -1 -II-
Cluster mode
Use training set
Supplied test set
Set...
Percentage split
%
Classes to clusters evaluation
(Nom) total
Store clusters for visualization
Ignore attributes
Start
Result list (right-click for options)
weka.gui.GenericObjectEditor
weka.clusterers.EM
About
Simple EM (expectation maximisation) class.
More
Capabilities
debug
False
displayModelInOldFormat
False
doNotCheckCapabilities
False
maxIterations
100
maximumNumberOfClusters
-1
minLogLikelihoodImprovementCV
1.0E-6
minLogLikelihoodImprovementIterating
1.0E-6
minStdDev
1.0E-6
numClusters
-1
numExecutionSlots
1
numFolds
10
numKMeansRuns
10
seed
10
Open...
Save...
OK
Cancel
Status
OK
Log
x 0





OBSERVATION:

Scheme: weka.clusterers.EM -I 100 -N -1 -X 10 -max -1 -ll-cv 1.0E-6 -ll-iter 1.0E-6 -M 1.0E-6 -K 10 -num-slots 1 -S 10

Relation: supermarket

Instances: 4627

Attributes: 217

[list of attributes omitted]

Test mode: evaluate on training data

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

EM

==

Number of clusters selected by cross validation: 1

Number of iterations performed: 1

Cluster

Attribute 0

(1)

==========

department1

t 4628

[total] 4628

department2

t 4628

[total] 4628……

Time taken to build model (full training data) : 19.92 seconds

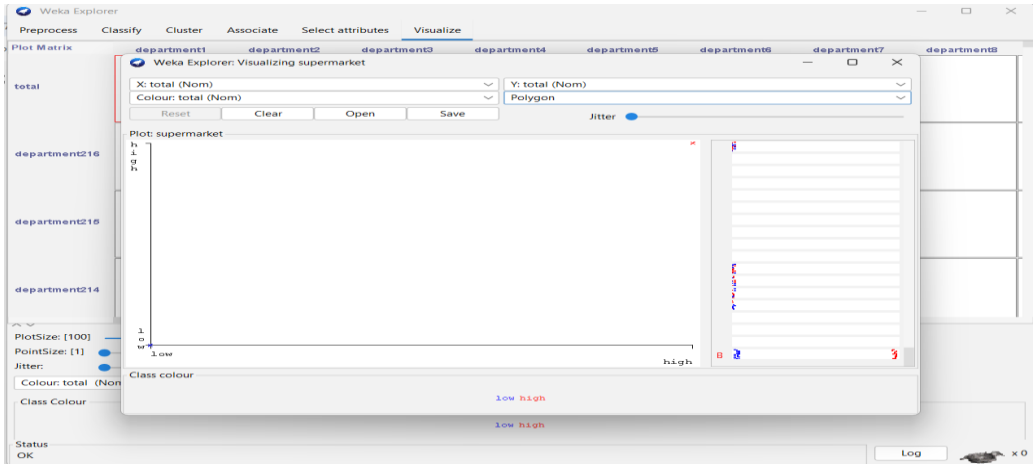
=== Model and evaluation on training set ===

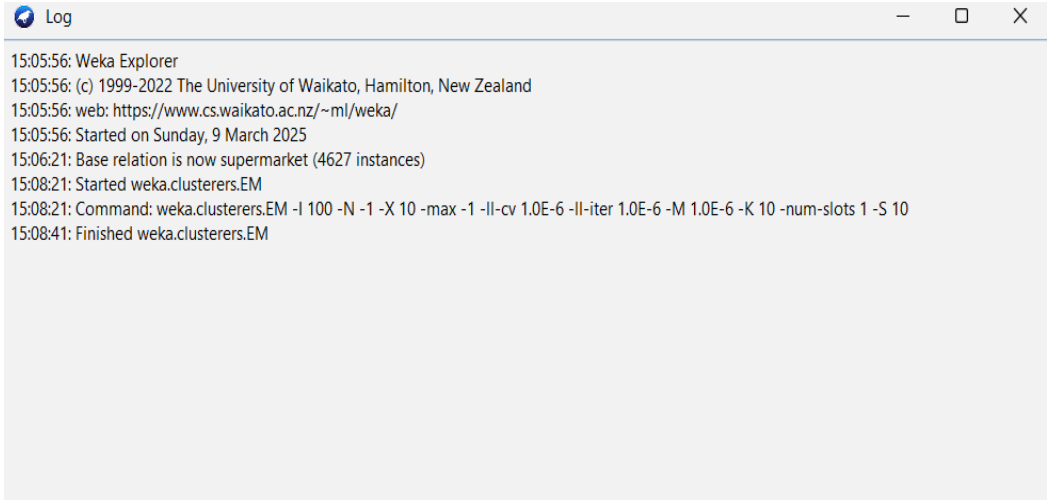
Clustered Instances

0 4627 (100%)

Log likelihood: -0.65505

PLOT:





RESULT:

Thus, the data analysis by the expectation maximization algorithm using weka has been analyzed and observed properly.