EXP 8: Demonstrate Classification, Clustering, Association using weka

Name: Vinit Prajapati
Class – TE-04
BATCH – B
Roll No – 40

Aim: To perform data preprocessing and demonstrate Classification, Clustering, and Association algorithms on datasets using the data mining tool WEKA.

Introduction (Theory):

WEKA (Waikato Environment for Knowledge Analysis) is an open-source data mining software that provides a collection of machine learning algorithms for data analysis and predictive modeling. It supports tasks such as:

- Classification: Assigning data to predefined categories (e.g., Decision Tree, Naive Bayes).
- Clustering: Grouping similar data points without predefined labels (e.g., K-Means).
- Association: Discovering interesting relationships or patterns in data (e.g., Apriori Algorithm).

Before applying these techniques, data must go through preprocessing, which includes cleaning, transforming, and preparing the data into a suitable format. WEKA offers a GUI for users to easily apply and visualize these techniques on various datasets (usually in .arff or .csv format).

Procedure:

Process of Weka

1.Open Weka and click on KnowledgeFlow. Click on Data Source and select Arff Loader.

2.Right click on Arff Loader and configure,then click on browse and select the iris dataset.Click OK.

3.Click on Cluster and select SimpleKMeans.

4.Click on Evaluation and select Training SetMaker and then Cluster Performance Evaluator.

5.Click on Visualisation and select Text Viewer.

6.Now for linking the datasets right click on Arff Loader and select dataset to connect with Training SetMaker.

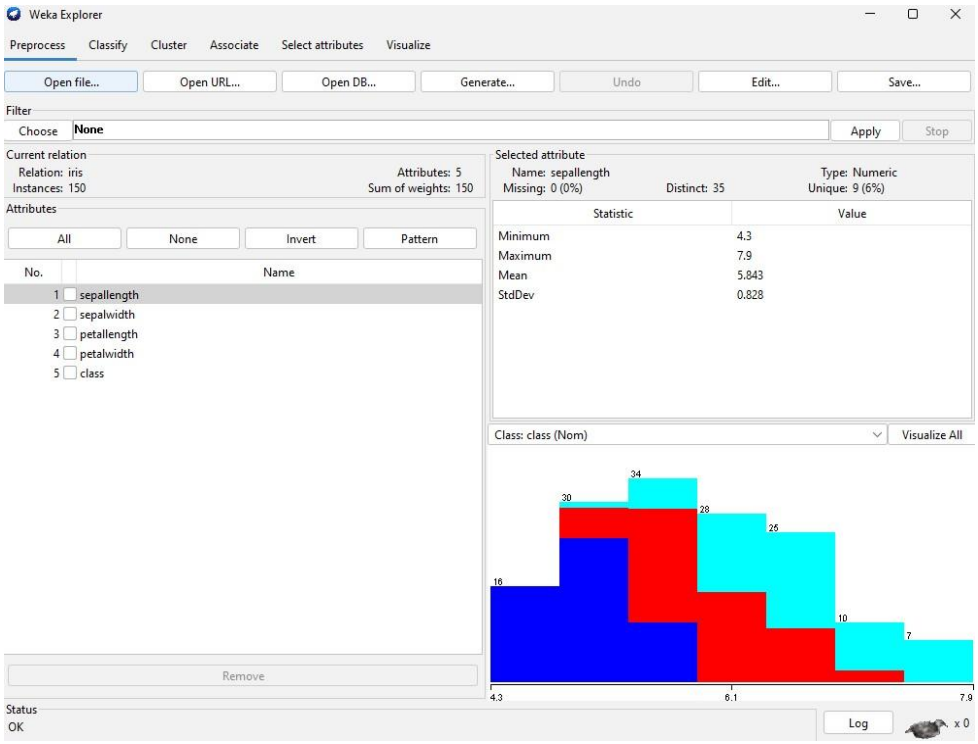
7.Right click on Training SetMaker and select training set to connect with SimpleKMeans.Now, right click on SimpleKMeans and select Batch Clusterer to connect with Clusterer Performance Evaluator.

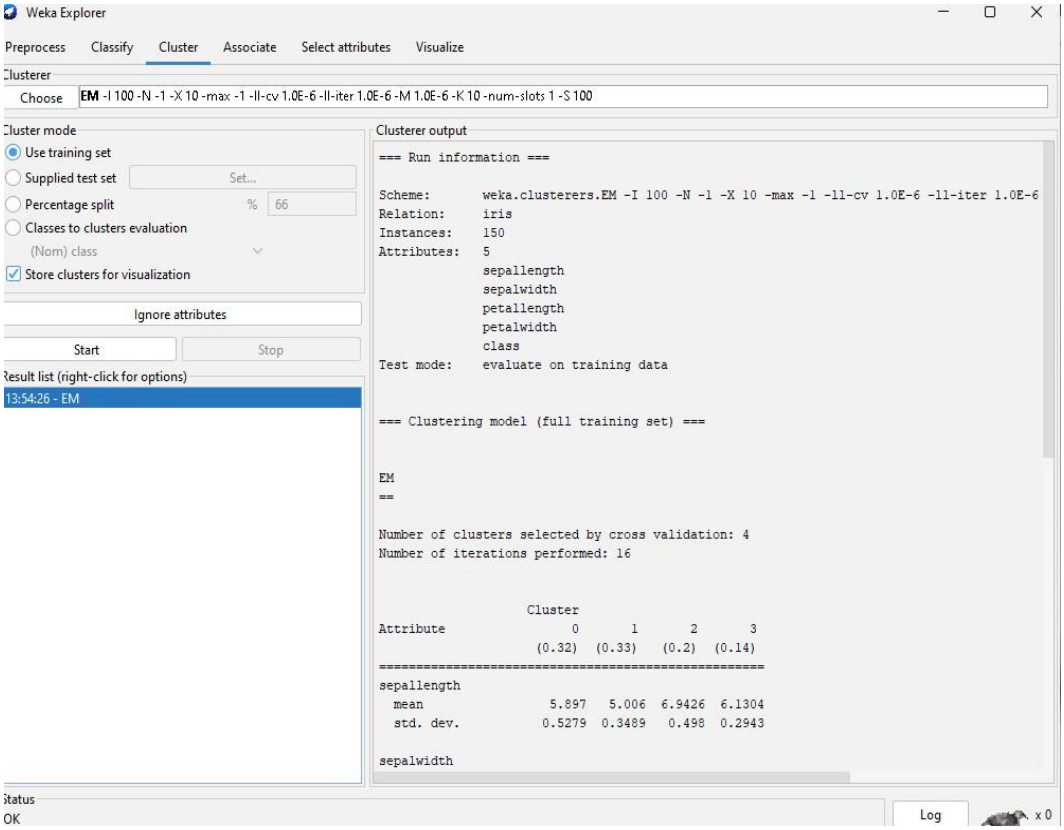
8.Right click on Clusterer Performance Evaluator and select text to connect with Text Viewer. Finally run and the status will be shown as Finished.

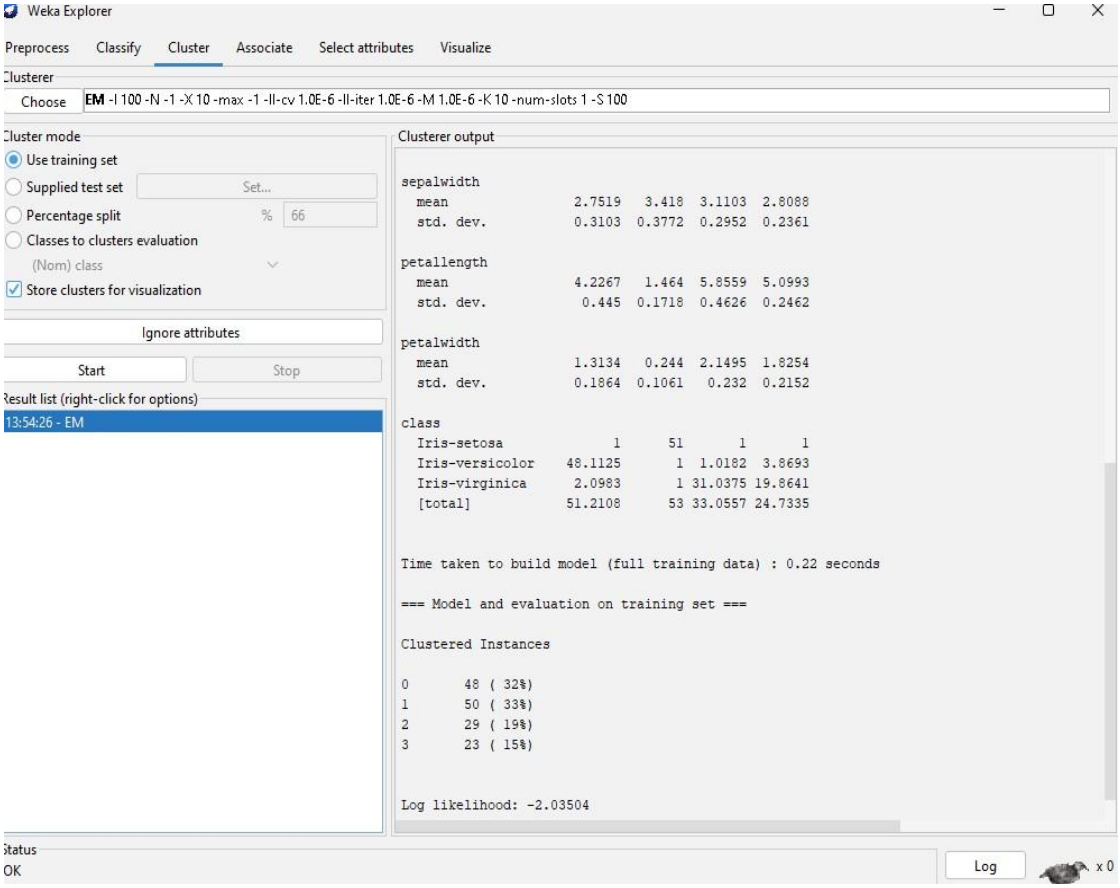
9.Now go to explorer in Weka and click on Open File and from Weka folder select data and click on iris dataset ann open it. Output for the preprocess and cluster will be displayed.

This process allows users to explore data, apply machine learning, and derive meaningful insights effectively

Implementation/Output snap shot:


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: iris
Instances: 150
Attributes: 5
Sum of weights: 150
Attributes
All
None
Invert
Pattern
No.
Name
1 sepallength
2 sepalwidth
3 petallength
4 petalwidth
5 class
Remove
Selected attribute
Name: sepallength
Missing: 0 (0%)
Distinct: 35
Type: Numeric
Unique: 9 (6%)
Statistic
Value
Minimum 4.3
Maximum 7.9
Mean 5.843
StdDev 0.828
Class: class (Nom)
Visualize All
34
30
28
26
16
10
7
4.3
6.1
7.9
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 -ll-cv 1.0E-6 -ll-iter 1.0E-6 -M 1.0E-6 -K 10 -num-slots 1 -S 100
Cluster mode
Use training set
Supplied test set
Set...
Percentage split
%
66
Classes to clusters evaluation
(Nom) class
Store clusters for visualization
Ignore attributes
Start
Stop
Result list (right-click for options)
13:54:26 - EM
Clusterer output
=== Run information ===
Scheme: weka.clusterers.EM -I 100 -N -1 -X 10 -max -1 -ll-cv 1.0E-6 -ll-iter 1.0E-6
Relation: iris
Instances: 150
Attributes: 5
sepallength
sepalwidth
petallength
petalwidth
class
Test mode: evaluate on training data
=== Clustering model (full training set) ===
EM
==
Number of clusters selected by cross validation: 4
Number of iterations performed: 16
Cluster
Attribute 0 1 2 3
(0.32) (0.33) (0.2) (0.14)
sepallength
mean 5.897 5.006 6.9426 6.1304
std. dev. 0.5279 0.3489 0.498 0.2943
sepalwidth
Status
OK
Log
x 0



Conclusion:

In this experiment, we successfully demonstrated the application of Classification, Clustering, and Association techniques using the WEKA data mining tool. WEKA provided a user-friendly interface for preprocessing data and applying various machine learning algorithms.

- Classification helped us categorize data based on labelled attributes.
- Clustering allowed us to group similar data instances without predefined labels.
- Association revealed interesting patterns and relationships within the dataset.

This experiment enhanced our understanding of key data mining concepts and gave us practical exposure to using WEKA for real-world data analysis tasks.

Review Questions:

1) Which WEKA panel allows you to load datasets and apply filters before modeling?

ANS: The Preprocess panel in WEKA is used to load datasets and apply various filters (like normalization, discretization, etc.) before performing any modeling or analysis.

2) What kind of learning does clustering fall under?
Ans: Clustering falls under Unsupervised Learning, where the algorithm tries to group data based on patterns without any predefined labels.

3) Which algorithm in WEKA is used to generate association rules?
Ans: The Apriori algorithm in WEKA is used to generate association rules that identify relationships between items in a dataset.

4) What evaluation metric in WEKA tells you how well your classification model is performing?
Ans: The Accuracy metric (often shown as "Correctly Classified Instances") indicates how well the classification model is performing on the dataset.

GITHUB LINK:-

https://github.com/Vin77777/DWM-EXP-8