



Vidyavardhini's College of Engineering and Technology

Department of Computer Engineering

Academic Year : 2023-24 (Odd Sem)

Experiment No.9
Clustering, Classification and Association Data Mining using WEKA tool
Date of Performance:
Date of Submission:



Vidyavardhini's College of Engineering and Technology

Department of Computer Engineering

Academic Year : 2023-24 (Odd Sem)

Aim: To implement clustering , classification and association data mining by using WEKA

Objective: Simulate K-Means Algorithm, Single Linkage Algorithm Decision tree induction and apriori algorithm by using WEKA

Theory:

WEKA, formally called Waikato Environment for Knowledge Learning, is a computer program that was developed at the University of Waikato in New Zealand for the purpose of identifying information from raw data gathered from agricultural domains. WEKA supports many different standard data mining tasks such as data preprocessing, classification, clustering, regression, visualization and feature selection. The basic premise of the application is to utilize a computer application that can be trained to perform machine learning capabilities and derive useful information in the form of trends and patterns. WEKA is an open source application that is freely available under the GNU general public license agreement. Originally written in C the WEKA application has been completely rewritten in Java and is compatible with almost every computing platform. It is user friendly with a graphical interface that

allows for quick set up and operation. WEKA operates on the predication that the user data is available as a flat file or relation, this means that each data object is described by a fixed number of attributes that usually are of a specific type, normal alpha-numeric or numeric values. The WEKA application allows novice users a tool to identify hidden information from database and file systems with simple to use options and visual interfaces.

1) K-Means Algorithm using WEKA

EXAMPLE:

Dataset: $D = \{1, 2, 3, 8, 9, 10, 25\}$

1. Randomly assign means $m1 = 3$ and $m2 = 10$

$k1 = \{1,2,3\}$ $k2 = \{8,9,10,25\}$

2. $m1 = 2$ and $m2 = 13$

$k1 = \{1,2,3\}$ $k2 = \{8,9,10,25\}$

WEKA Code:

@RELATION iris



Vidyavardhini's College of Engineering and Technology

Department of Computer Engineering

Academic Year : 2023-24 (Odd Sem)

@ATTRIBUTE x NUMERIC

@DATA

1

2

3

8

9

10

25



Vidyavardhini's College of Engineering and Technology

Department of Computer Engineering

Academic Year : 2023-24 (Odd Sem)

Vidyavardhini's College of Engineering and Technology

Department of Computer Engineering



2) Decision Tree Induction using WEKA

A decision tree is a flowchart like tree structure, where each internal node(non-leaf node) denotes a test on an attribute,each branch represents an outcome of the test,and each leaf node (or terminal node) holds a class label. The topmost node in a tree is the root node

Example:-

Outlook	Temperature	Humidity	Windy	Class
sunny	hot	high	false	N
sunny	hot	high	true	N
overcast	hot	high	false	P
rain	mild	high	false	P
rain	cool	normal	false	P
rain	cool	normal	true	N
overcast	cool	normal	true	P
sunny	mild	high	false	N
sunny	cool	normal	false	P
rain	mild	normal	false	P
sunny	mild	normal	true	P
overcast	mild	high	true	P
overcast	hot	normal	false	P
rain	mild	high	true	N

Output:-



Vidyavardhini's College of Engineering and Technology

Department of Computer Engineering

Academic Year : 2023-24 (Odd Sem)

3) Apriori Algorithm using WEKA

In this current world, globalization is the main feature of any environment. Everyone has to be update, fast and forward and information is the main

element for it. For survival in this world it's the basic need to use and to store the information means to prepare a proper database or dataset to analyze. Using and storing the database is not an issue, but finding the relevant dataset or to analyze the meaningful dataset for a particular aspect, from the junkyard of the database is very big problem in analysis of a specific part of the database. To solve this problem the concept of data mining is used to abstracts the desirable information. Useful information from the large databases has been extracted in the form of the association rules. There are many algorithms have been developed to extract the association rules from the large databases. Apriori algorithm is the most popular algorithm to extract the association rules from the databases.

TID	Items
1	A,B,C,D,G,H
2	A,B,C,D,E,F,H
3	B,C,D,E,H
4	B,E,G,H
5	A,B,D,E,G,H
6	A,C,F,G,H
7	B,D,E,G,H
8	A,C,D,E,G,H
9	B,C,D,E,H
10	A,C,E,F,H
11	C,E,H
12	A,D,E,F,H



Vidyavardhini's College of Engineering and Technology

Department of Computer Engineering

Academic Year : 2023-24 (Odd Sem)

		15	A,B,E,F,H
13	B,C,E,F		
14	A,B,C,F		

Example

Output



Vidyavardhini's College of Engineering and Technology

Department of Computer Engineering

Academic Year : 2023-24 (Odd Sem)



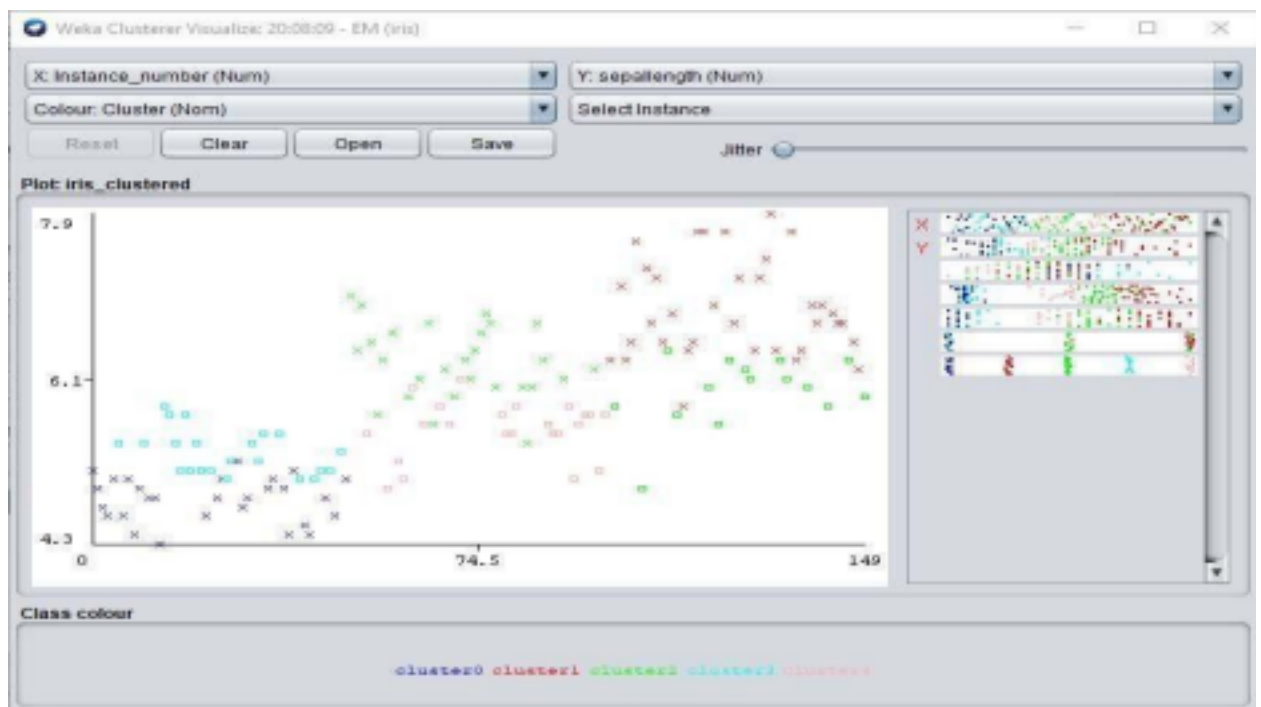
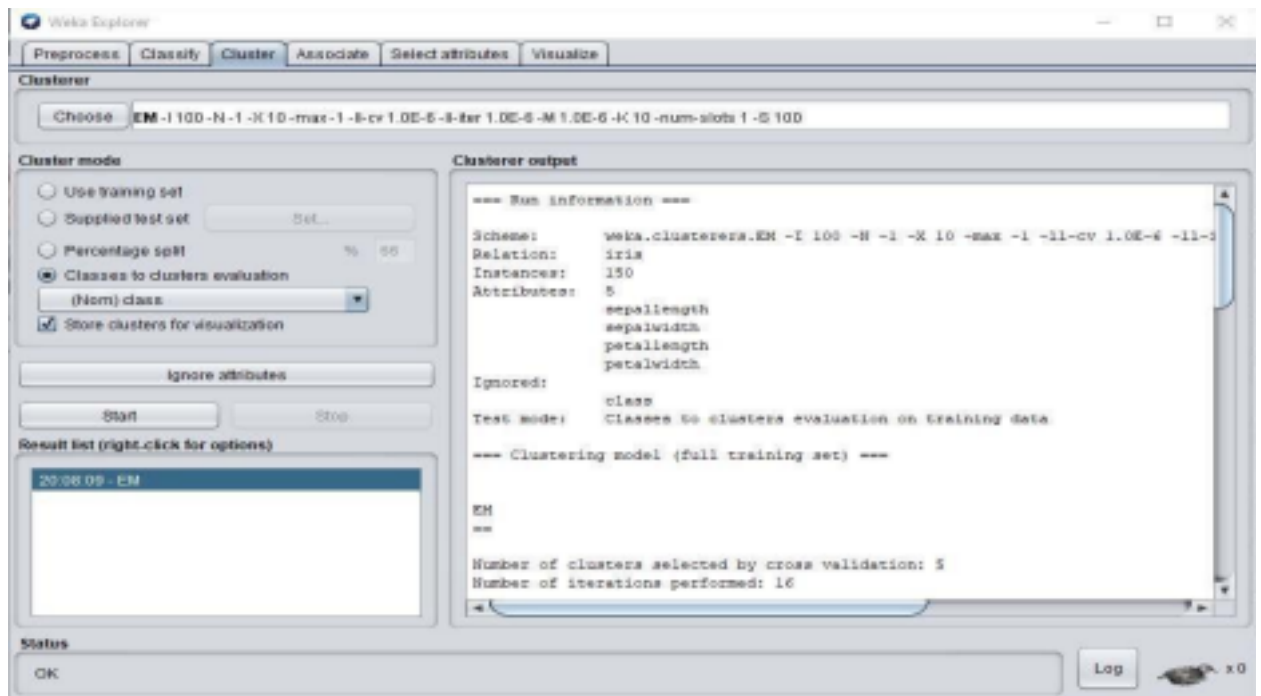
Code and output:

Vidyavardhini's College of Engineering and Technology

Department of Computer Engineering

Academic Year : 2023-24 (Odd Sem)

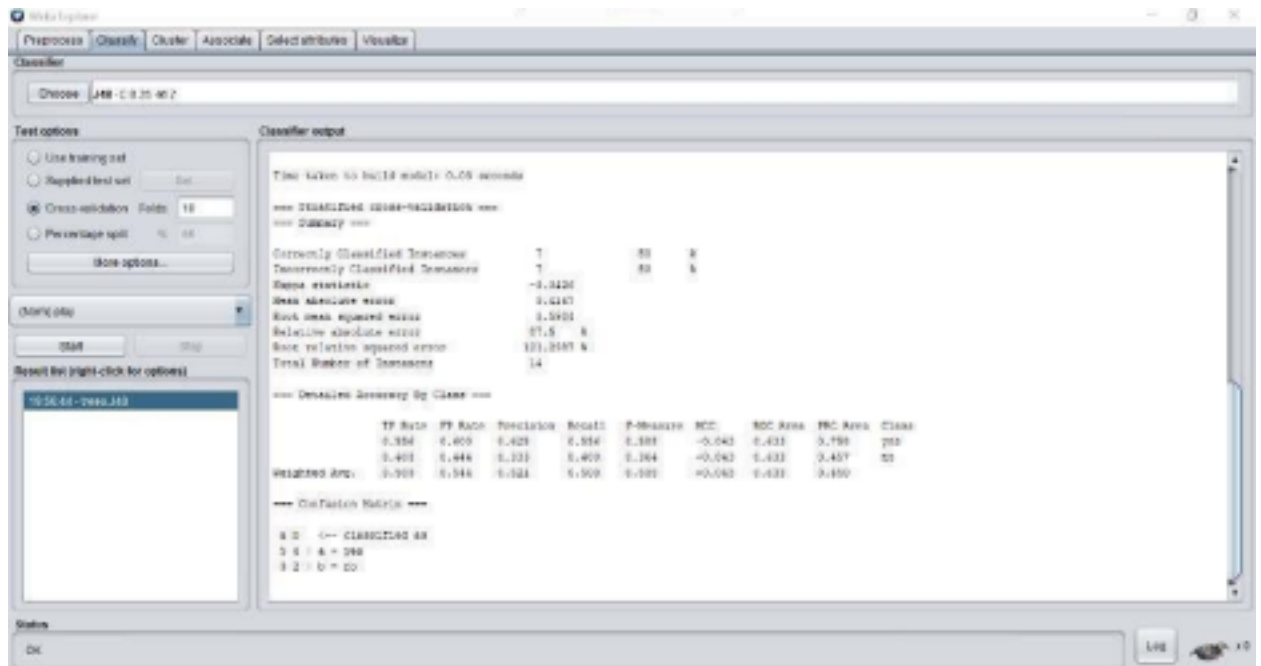




Vidyavardhini's College of Engineering and Technology

Department of Computer Engineering

Academic Year : 2023-24 (Odd Sem)



Vidyavardhini's College of Engineering and Technology

Department of Computer Engineering

Academic Year : 2023-24 (Odd Sem)



Vidyavardhini's College of Engineering and Technology

Department of Computer Engineering

Academic Year : 2023-24 (Odd Sem)

Conclusion:WEKA (Waikato Environment for Knowledge Analysis) is a popular open-source machine learning software that provides a user-friendly interface for various data mining and machine learning tasks.WEKA offers a graphical user interface (GUI) that allows users, even those without extensive programming or data science experience, to perform a wide range of machine learning tasks. WEKA provides a comprehensive set of preprocessing tools for data cleaning, transformation, and feature selection. WEKA simplifies the process of selecting machine learning algorithms and evaluating their performance.WEKA provides visualization tools that make it easier to understand and interpret the results of machine learning experiments. Once a model is trained, WEKA allows users to save it for future use. WEKA often provides parameter tuning options for algorithms, making it easier to fine-tune model performance without extensive manual experimentation. In summary, WEKA simplifies the machine learning process by offering a user-friendly interface, a rich set of preprocessing and modeling tools, visualization capabilities, and automated techniques for evaluation and parameter tuning. The outputs it generates, including cleaned datasets, performance metrics, and visualizations, assist users in making informed decisions throughout the entire machine learning workflow. This makes WEKA a valuable tool for both beginners and experienced data scientists in their data analysis and modeling endeavors.