Name : Suvidhi V. Pareek

Class / Div : TE / 03

Roll No : 09

| Experiment No.9 |
| --- |
| Clustering, Classification and Association Data Mining using WEKA tool |
| Date of Performance: 24/09/24 |
| Date of Submission: 01/10/24 |

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

**Objective:** Simulate K-Means Algorithm, Single Linkage AlgorithmDecision 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}

1. m1 = 2 and m2 = 13

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

**WEKA Code:**

@RELATION iris

@ATTRIBUTE x NUMERIC

@DATA

1

2

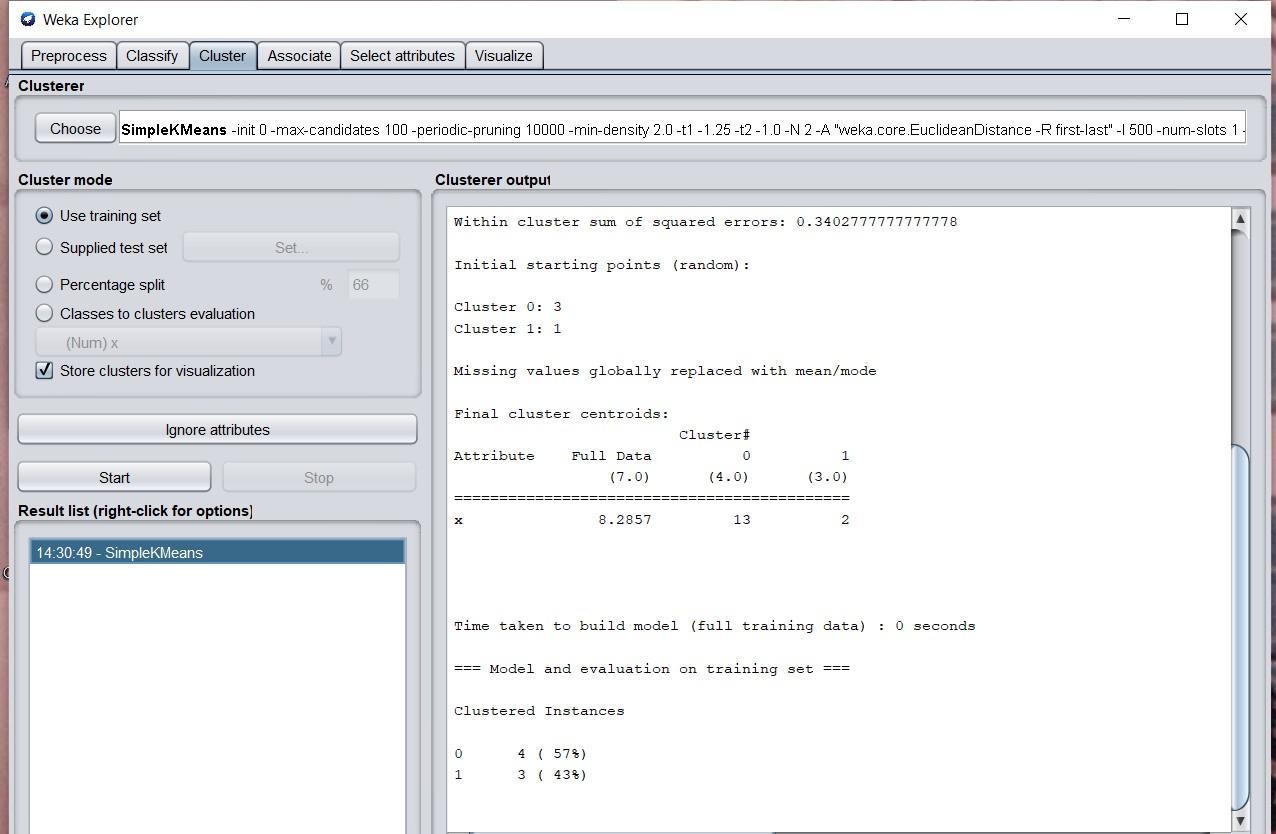
3

8

9

10

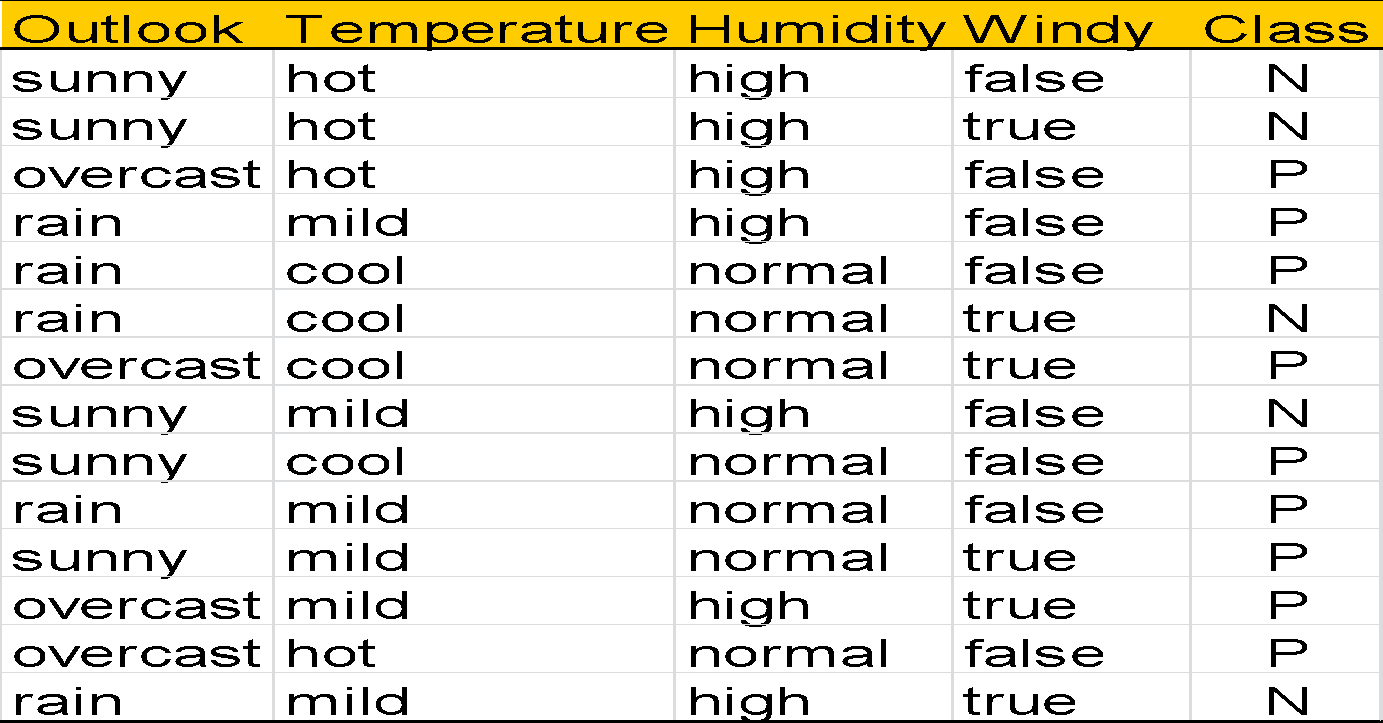
25

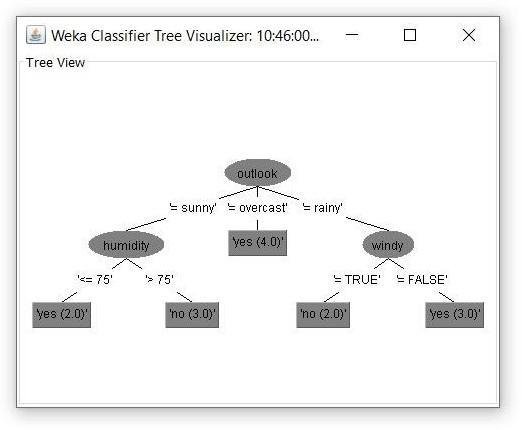


1. **Decision Tree Induction using WEKA**

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:-



Output:-

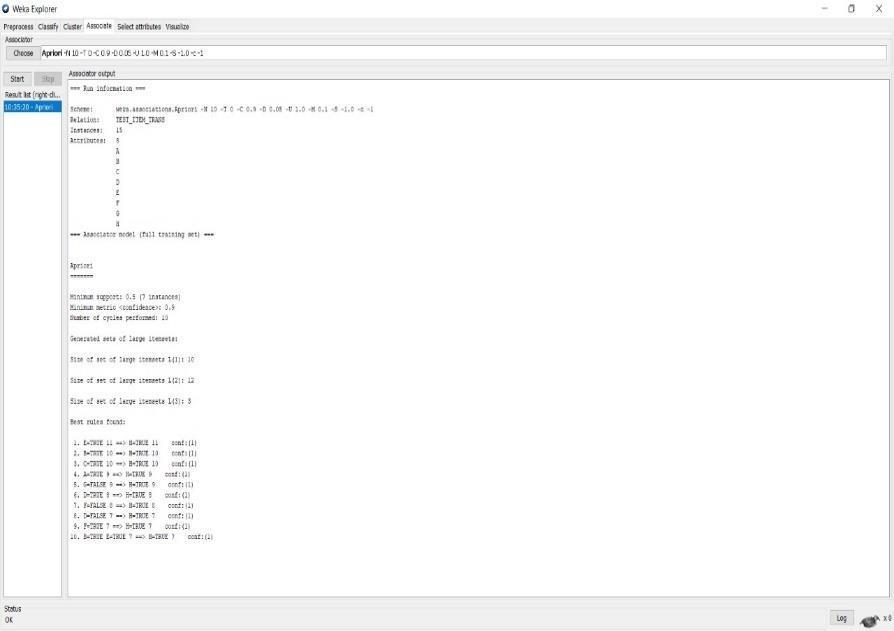
1. **Apriori Algorithm using WEKA**

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.

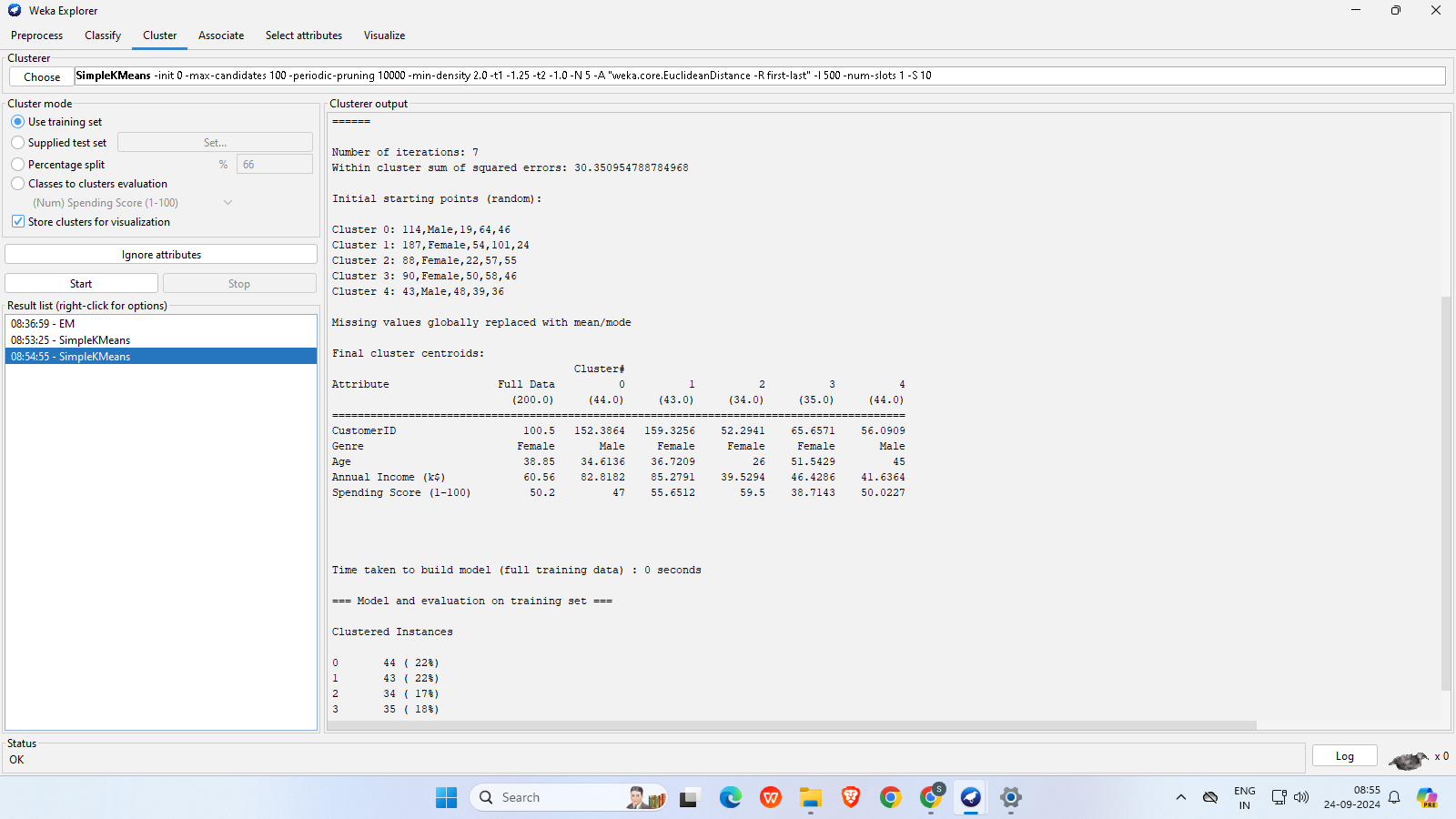
**Example**

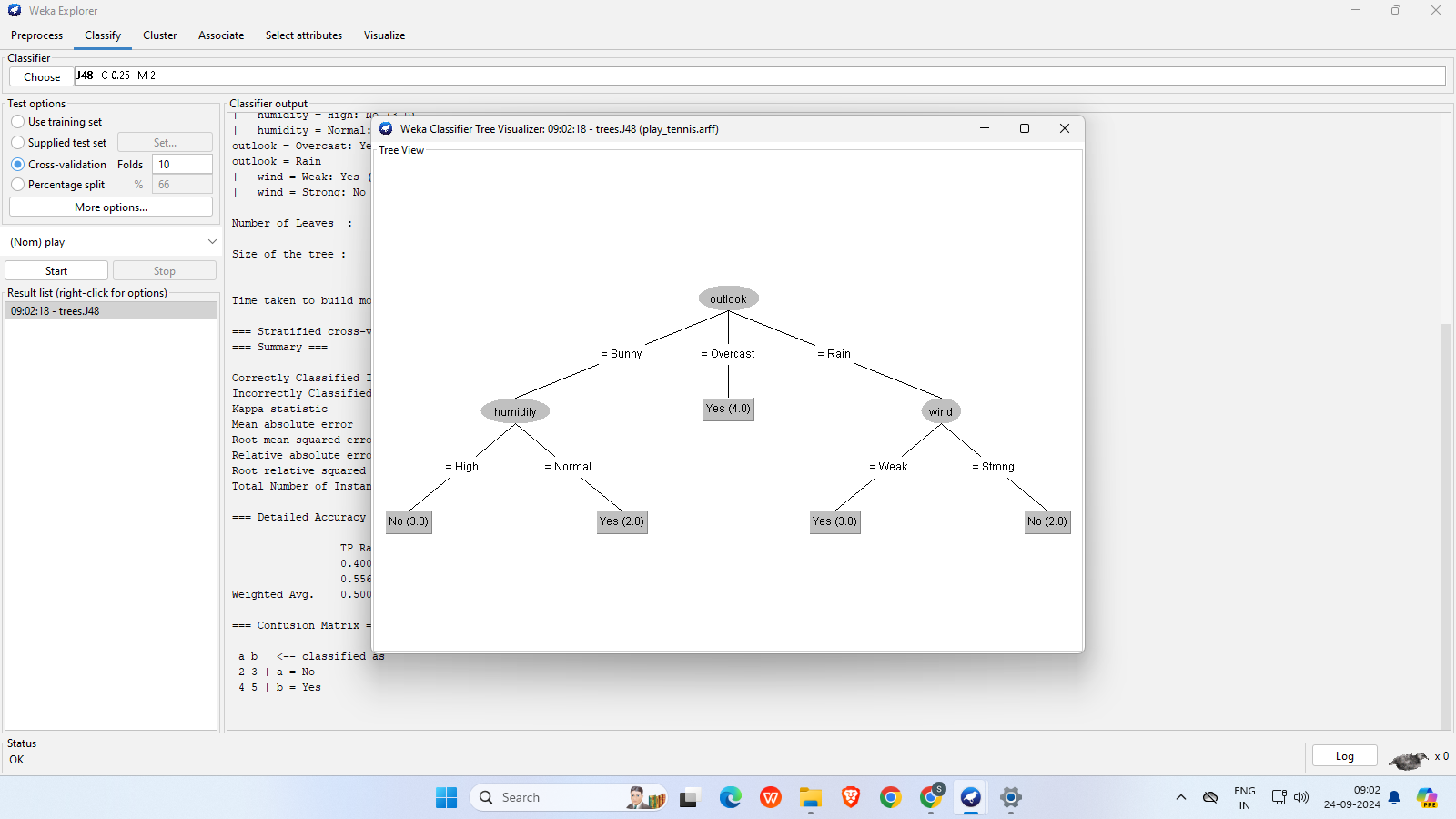
| TID | Items |
| --- | --- |
| 1 | A,B,C,D,G,H |
| 2 | A,B,C,D,E,F,H |
| 3 | B,C,D,E,H |
| 4 | B,E,G,H |
| 5 | A,B,D,E,G,H |
| 6 | A,C,F,G,H |
| 7 | B,D,E,G,H |
| 8 | A,C,D,E,G,H |
| 9 | B,C,D,E,H |
| 10 | A,C,E,F,H |
| 11 | C,E,H |
| 12 | A,D,E,F,H |
| 13 | B,C,E,F,H |
| 14 | A,B,C,F,H |
| 15 | A,B,E,F,H |

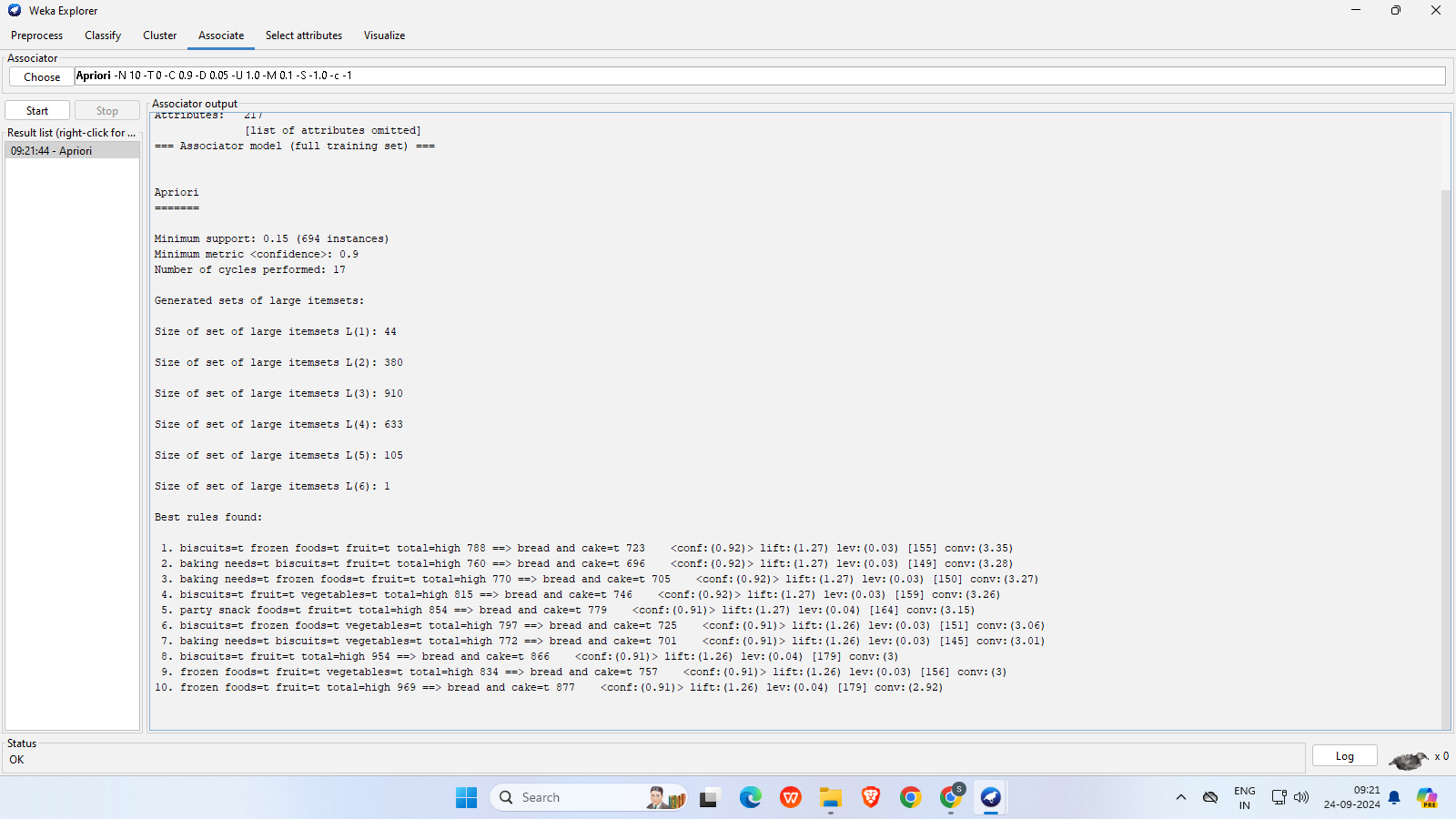




**Code and output**:







**Conclusion**: Comment on the outputs that gets generated and how WEKA simplifies the process. WEKA simplifies clustering, classification, and association rule mining by providing an easy-to-use interface for implementing algorithms like K-Means, Decision Trees, and Apriori, making data analysis accessible without extensive coding.