**Name:** Soham Belurgikar

**Roll No.:** 2019130006

**Course:** **DA (Data Analytics)**

**Experiment No.:** 5

**Name of the Experiment:** Apriori Algorithm and Association rule mining with WEKA

**Objective:** Apply Apriori Algorithm to given dataset

Association Rule Mining with WEKA

**Implementation:**

**Task:** <https://colab.research.google.com/drive/1sANAFvw-k3ia1lqGNKGDzb5E7ZDdy0IF?usp=sharing>

**Exercise 1**:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Transaction | A | B | C | D | E | K |
| T1 | 1 | 1 | 0 | 1 | 0 | 1 |
| T2 | 1 | 1 | 1 | 1 | 1 | 0 |
| T3 | 1 | 1 | 1 | 0 | 1 | 0 |
| T4 | 1 | 1 | 0 | 1 | 0 | 0 |

Minimum support of 60% => item has to occur in at least 3 transactions

Item sets containing 1 item:

A 4, 100%

B 4, 100%

C 2, 50%

D 3, 75%

E 2, 50%

K 1, 25%

Item sets containing 2 items:

We only take the item sets from the previous phase whose support is 60% or more.

A B 4, 100%

A D 3, 75%

B D 3, 75%

Item sets containing 3 items:

We only take the item sets from the previous phase whose support is 60% or more.

A B D 3

Form the rules and calculate their confidence (c). We only take the item sets from the previous phases whose support is 60% or more.

Rules:

A -> B P(B|A) = |B∩A| / |A| = 4/4, c: 100%

B -> A c: 100%

A -> D c: 75%

D -> A c: 100%

B -> D c: 75%

D -> B c: 100%

AB -> D c: 75%

D -> AB c: 100%

AD -> B c: 100%

B - > AD c: 75%

BD -> A c: 100%

A -> BD c: 75%

The rules with a confidence measure of 75% are pruned, and we are left with the following rule set:

A -> B

B -> A

D -> A

D -> B

D -> AB

AD-> B

DB-> A

**Exercise 2**:

Input file: test.arff

Text

Description automatically generated

Weka configuration file:

Graphical user interface, application

Description automatically generated

Output:

A picture containing text

Description automatically generated

This output is consistent with the manual solution in Exercise 1. The file format used was .arff.

**Exercise 3**:

A picture containing graphical user interface

Description automatically generated

**Exercise 4**:

Graphical user interface, table

Description automatically generated

Graphical user interface, application

Description automatically generated

**Exercise 5**:

Graphical user interface, application

Description automatically generated

**Conclusion:**

* Because it relies on prior knowledge about frequent itemset properties, the algorithm is called Apriori. We employ an iterative strategy, often known as level-wise search, in which we use k-frequent itemsets to find k+1 itemsets.
* Association rule mining finds interesting associations and relationships among large sets of data items. This rule shows how frequently a itemset occurs in a transaction.
* In exercise 4, most of the generated rules have similar confidence and lift levels. None of the rules in the default output involve Class = republican. This is because of the lack of significant number of members belonging to republican party in the chosen dataset.