**﷽**

**“A COMPREHENSIVE REPORT”**

**of**

**“2nd Assignment of Data Mining”**

|  |
| --- |
| **PROGRAM** |
| **“Master of Computer Science”** |
| **RESOURCE PERSON** |
| **“Miss Rabia Naseer Rao”** |
| **PREPARED BY** |
| **Syed Shahzad Mubashar (ID#F2019027030)** |
| **Muhammad Muzamil (ID#F2019027042)** |
| **Abdullah Shahid (ID#F2019027004)** |

“School of Professional Advancement, University of Management and Technology, C-II Johar Town, Lahore.”

**“TASK”**

**“TASK FOR THIS ASSIGNMENT IS TO IMPLEMENT AND EVALUATE THE APRIORI BASED ALGORITHM FOR FREQUENT ITEM SETS MINING.”**

**First of all, we have started with the introduction of Apriori Algorithm. Then we have made a code which we will use for the implementation of our Assignment. Our code has 14 sub functions and 2 Main Functions which are used to find the Frequent Item Sets, Their rules and it will find the Support, Confidence and Lift as well. Our Main Functions are as under:**

* **My\_apriory.**
* **Generate\_role\_and\_measure\_support\_confidence\_lift.**

**“1. My\_Apriory”**

**This function will take two parameters and return 2-D list of frequent items set using the approach of Hash Table Algorithm that is Apriory with Hash Table Algorithm to find Frequent Items set.**

* **All\_Items\_List:**

**A 2-D List for all the transactions which will remove the NAN Values (Null values/empty values) from our dataset.**

* **Support:**

**This is an integer parameter refers to minimum support value used to find Frequent Item Sets.**

**“2. GENERATING RULES AND MEASURING SUPPORT AND CONFIDENCE”**

**This and save their output in a text file which we will give as a parameter and it will return all possible rules with their support, confidence and lift value. This functions have 3 parameters as under:**

* **Frequent\_Items.**
* **All\_Items\_List.**
* **File\_Name**

**“Frequent Items”**

**This parameter takes 2-D list of Frequent Items Set returned by My\_Apriory function.**

**“All\_Items\_List”**

**This is a 2-D list of all transactions which we load from our Dataset.**

**After that we have applied all the functions on 3 different data sets respectively.**

**“File Name”**

**This parameter takes the string for the name of that file which will save the all output.**

**“First Dataset”**

**Our First data set is a dummy dataset prepared by ourselves so that we can easily understand how the implementation works. That is this data set is only used for understanding/testing purposes. As the set is small so it is possible to see the set in this report:**

|  |  |
| --- | --- |
| **INSTANCE** | **ITEMS** |
| **I1** | d,e,a |
| **I2** | d,b,e,f |
| **I3** | d,e,c |
| **I4** | e,c,a,d |
| **I5** | b,g,i,f,a,c |
| **I6** | f,g,e,c,a |
| **I7** | b,g,i,f,a,c,e,d |

**We start with Data understanding that is reading the dataset, then we** Execute Apriory Algorithm with Minimum Support equal to 2. Then we find the association rules after that we calculate Support, Confidence and Lift.

After that we execute it with Min\_Support = 3 and repeat the whole phase again. At the end we have shown the association rules and the values of support, confidence and lift in a table.

**“Second Dataset”**

**Our First data working has been complete then we move towards the Second Data Set which is also a dummy data set having 50 Instance.** At the end we have shown the association rules and the values of support, confidence and lift in a table. We have repeated the work with three different values of Min\_Support that is for 15, 10 and 8 for building our concepts strongly.

**“Third Dataset”**

**Our Third data set is** a real world dataset which belongs to a market. Which is downloaded from link given below:

# https://stackabuse.com/association-rule-mining-via-apriori-algorithm-in-python

This dataset has more than 7000 instances and more than100 items. This is used for assuring purposes that we can also apply this algorithm on a larger dataset. We pass this dataset to our Apriory Algorithm and also pass the min\_support value to find the frequent items, pass that frequent item list to second function **"Generate\_Role\_And\_Measure\_Support\_Confidence\_Lift"** to find the rules and calculation of support, confidence and lift according to that rules. We have repeated the work with three different values of Min\_Support that is for 20, 30 and 60 for building our concepts strongly.

**\*\_\_\_\_\_\_\_\_\_\_THE END\_\_\_\_\_\_\_\_\***