**Experiment 7: Association Rule Analysis using Python**

**Objective** :To implement association rule analysis using Python.

**Time Required** : 3 hrs

**Programming Language** : Python

**Software Required** : Anaconda

**Introduction**

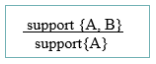
Association Rule Analysis finds interesting associations and relationships among large sets of data items. This rule shows how frequently an itemset occurs in a transaction. A typical example is Market Based Analysis. Market Based Analysis is one of the key techniques used by large relations to show associations between items. It allows retailers to identify relationships between the items that people buy together frequently.

**Apriori Algorithm:**

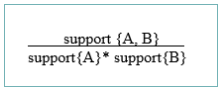
Apriori algorithm is used for finding frequent item-sets in a dataset for Boolean association rule. Name of the algorithm is Apriori because it uses prior knowledge of frequent itemset properties. We apply an iterative approach or level-wise search where k-frequent item-sets are used to find k+1 item-sets. To improve the efficiency of level-wise generation of frequent item-sets, an important property is used called Apriori property which helps by reducing the search space.

The output of the apriori algorithm is the generation of association rules. This can be done by using some measures called support, confidence, and lift. Now let’s understand each term.

**Support:** It is calculated by dividing the number of transactions having the item by the total number of transactions.

**Confidence:** It is the measure of trustworthiness and can be calculated using the below formula.

Conf(A => B)=

**Lift:** It is the probability of purchasing B when A is sold. It can be calculated by using the below formula.

Lift(A => B)=

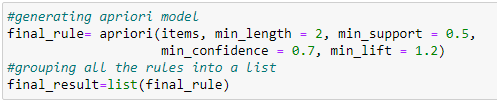
**Lift(A => B) =1 :** There is no relation between A and B.  
**Lift(A => B)> 1:** There is a positive relation between the item set . It means, when product A is bought, it is more likely that B is also bought.  
**Lift(A => B)< 1:** There is a negative relation between the items. It means, if product A is bought, it is less likely that B is also bought.

**Implementing Apriori With Python**

First, we need to install the apyori package. To install the package, open the terminal or command prompt, type in and enter the following command:

***pip install apyori***

1. **Import the libraries (apriori from apyori , numpy and pandas) and load the dataset.**
2. Convert the data frame into lists. The algorithm in the apyori package is implemented in such a way that the input to the algorithm is a list of lists rather than a data frame. So, we need to convert the data into a list of lists.
3. Generate the Apriori Model



Where,

* min\_support: The minimum support of relations (float)
* min\_confidence: The minimum confidence of relations (float)
* min\_lift: The minimum lift of relations (float)
* min\_length: The minimum number of items in a rule
* max\_length: The maximum number of items in a rule

The optimum values for min\_support,min\_confidence and min\_lift arguments can be set by trying out different values and checking the association rules whether the arguments produced a valid association between items or not.

1. Display the final rules: The Apriori algorithm automatically sorts the associations’ rules based on relevance, thus the topmost rule has the highest relevance compared to the other rules returned by the algorithm. Rule one is the most relevant rule that the algorithm identified from the given dataset.

**TASK:**

1. Load and print the dataset. What do the rows and columns represent in dataset? What does the value ‘NaN’ in dataset indicate?

**Note:** Your dataset must not contain any header. To ensure this, set ‘header=None’ in a script where you load and read your dataset. Call the head() function. In the output you will see that the first line is treated as a record instead of header.

1. Run Apriori Algorithm on given dataset when:

min\_length=2, min\_support = 0.0045, min\_confidence = 20%, min\_lift = 3.

1. Print all the association rules generated for above dataset. Also, count the number of association rules.
2. Print the first association rule from the given dataset. Also, explain what does it signify?