ASSIGNMENT 6

Title: Association Rule Learning

Problem Statement:

Downloading the Market Basket Optimization dataset It contains a total of 7501 transaction records where each record consists of the list of items sold in one transaction. Find association rules betn. items.

a. Data Preprocessing
b. Generate list of transactions from dataset

c. Train apriori algo. on dataset d. Yisualize list of rules

ctive:

It will # help us understand and implement Apriori algorithm.

S/W Packages & H/W apparatus: Linux os: Ubuntu / Windows, Jupyter notebook Pentium IV 1.7 GHz, 128 MB Ram, 40 GB HDD, etc.

The Apriori algo. is used for the purpose of associative rule mining. A.R.M is a technique to identify frequent patterns and association among items cg. By finding correlations and association before diff. items and that customers place in their basket, patterns can be derived.

Apriori Algorithm:

It assumes that any subset of a frequent item-

set must be trequent.

Say a transaction containing (wine, chips, bread)

also contains (wine, bread). So acc. to principle of

Apriori, if (wine, chips, bread) is frequent, then

(wine, bread) must also be frequent.

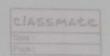
The key concept in this algo is that it assumes all its subsets of a frequent itemset to be

Decreated Similarly 2 and infrequent itemset all

frequent. Similarly, for any infrequent itemset, all its superset must also be infrequent.

Here is a dataset containing of six transac. tions in an hour. Each transaction is a combination of 0's and 1's where 0 represents
the absence of an item and 1 represents
the presence of it.

Transaction ID	Wine	Chips	Brea d	Milk	-
1	1	1	1	1	
2	1	0	1	1	
3	0	0	1	1	
4	0	1	0	D	
5	1	1	1	1	
6	1	1	٥	1	



We can find multiple rules from this scenario.

In a transaction of wine, chips & bread, it wine & chips are brought, then so is bread.

(wine, chips) > (bread).

In order to select the interesting rules out of multiple possible rules from this small business scenario, we will be using the following measures: · Support · List · Confidence · Conviction

1 Support It's nothing but the ratio of the no. of transactions in which item & appears to the fotal no. of transactions.

i.e. Support (wine) = No. of transactions containing wine
Total no. of transactions

Support (wine) = 4/6 = 0.6667

1 Confidence Confidence (x = 7 y) signifies the likelihood of y being purchased when x is purchased.

i.e. Conf. ((wine, chips) = (bread)) = Support (wine, c, 8) support (W,C)

 $(\text{onf.}((w, c) \Rightarrow (B)) = \frac{2/6}{3/6} =$ 0.6667 3 Lift

Lift $(x \Rightarrow y)$ is nothing but the 'interestingness' or the likelihood of y being purchased when x is sold. Unlike confidence $(x \Rightarrow y)$, this method takes into account, popularity of item y.

Lift($\{W, (Y \Rightarrow \{B\}) = Support(W, (B))\}$ Support(W, (Y)) = 2/6 = 13/6 * 4/6

- · Lift (x => y) = 1 means there's no correlation within itemset. · Lift (x => y) 7 1 :- there's a +ve correlation
- · Lift (x=>y) < 1 :- there's a -ve correlation

Conclusion:
Understood and successfully implemented the association rule learning algo on given dataset.