Assigment 3

Title : Apriori algorithm

Problem statement :
Apply apriori algorithm to find frequently occuring items from given data and generate strong association rules using support and confidence thresholds.

Data : Market Basket Analysis

Software and Hardware requirements :

Jupyter notebook, 2GB RAM, 500 GB HDD.

Objective : Apriori functions use to implement apriori algorithm

Outcome : Mining of frequent itemsets and create or generate relevant association rules

Theory :
With the quick growth in e-commerce applications, there is an accumulation of vast quantity of data in months not in years.
Data minning also known as Knowledge Discovery in Databases (KDD) to find

anomalies, correlations, patterns and trends to predict outcomes.

Apriori algorithm is a classical algorithm in data minning. It is used for mining frequent itemsets and relevant association rules. It is divised to operate on a database containing a lot of transactions.

The significant components compromise the apriori algorithm.

- → Support
- → Confidence
- → Lift.

Support : Support is the default popularity of any item. You calculate support as a quotient of the division of the number of transactions containing that item by total ~~numbe~~ number of transactions.

$$\text{Support} = \frac{\text{Transaction involving Jam}}{\text{Total transactions}}$$

Confidence :

It is the likelyhood that customers bought both bread and jam. Dividing no of transactions that include both bread and jam by total number of transactions will give the confidence figure.

$$\text{Confidence} = \frac{\text{Transactions having both bread \& jam}}{\text{Total transactions having jam}}$$

Lift :

It is the increase in the ratio of sale of bread when you sell jam

$$\text{Lift} = \frac{\text{confidence (Jam - Bread)}}{\text{support (Jam)}}$$

It says chance of customer buying both jam and bread together is 5 times more than jam alone.

If lift value is less than 1, customers are unlikely to buy both items together.

Algorithm :

1) Apriori (Pharm GKB, ϵ)

2) $L_1 \leftarrow$ (frequent genes in drug class for)
At

3) $K \leftarrow 2$

4) while $L_{k-1} \neq \phi$

5) $C_k \leftarrow \{a \cup \{b\} \mid a \in L_{k-1} \wedge b \in \cup L_{k-1} \wedge b \notin a\}$

	Test Cases	Expected O/P	Actual O/P
1	Install apriori library	success	success
2	Preprocess data	success	success
3	Training apriori on dataset	success	success
4	Visualize the results	success	success
5	Create association rules	success	success

Conclusion : Thus we implemented apriori algorithm using python for given transaction input.