Experiment No. 09 Date :

* Aim : Implantation of Association rule mining - Ariori Algorithm in Python.

* Theory :

The Apriori algorithm is an unsupervised machine algorithm used for association rule learning. Association rule learning is a data mining technique that identifies frequent patterns, connection and dependencies among different groups of items called itemsets in data. Aprogri algorithm refers to the algorithm refers to the algorithm which is used to calculate the association rules between objects. It means how two or more objects are related to one other. In the words, we can say that the apriori algorithm is an association rule learning that analyzes that people who bought product A also bought product B.

Apriori algorithm is given by R. Agrawal and R. Srikant in 1994 for finding frequent itemsets 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 itemsets are used to find k+1 intemsets.

To improve the efficiency of level-wise generation of frequent itemsets, an important property is used

Page No:

FOUNDATION

Experiment No. Date :

* Confidence :-

Confidence refers to the possibility that the customers bought both biscuits and chocolates together. So, you need to divide the number of transactions that comprise both biscuits and chocolates by the total number of transactions to get the confidence.

Hence,

Confidence (A → B) = (Transactions containing both (A and B)) / (Transactions containing A)

= 200 / 400

= 50 percent.

It means that 50 percent of customers who bought biscuits also bought chocolates.

* Lift :

Consider the above example ; lift refers to the increases in the ratio of the sale of chocolates when you sell biscuits. The mathematical equations of lift are given below.

Lift (A → B) = (Confidence (A → B)) / (Support (B))

= 50 / 10

= 5

It means that the probability of people buying both biscuits alone. If the lift value is below one, it requires that the people are unlikely to buy both the items together. Larger the value, the better the combination.

Page No:

called Apriori property which helps by reducing the search space.

Apriori Property -

All non-empty subset of frequent itemset must be frequent. The key concept of Apriori algorithm is its anti-monotonicity of support measure. Apriori assumes that

- All subsets of a frequent itemset must be frequent (Apriori property).
- If an itemset is infrequent, all its supersets will be infrequent.

* Components of Apriori Algorithm:

The given three components comprise the aprior algorithm.

1. Support
2. Confidence
3. Lift

* Support:

Support refers to the default popularity of any product. You find the support as a quotient of the division of the number of transactions comprising that product by the total number of transactions. Hence, we get

Support (B) = (Transactions containing (B))/(Total Transactions)
= 400/4000
= 10 percent.