



Assignment - 07

myCOMPANION

43304

* Aim → Create association rules for market basket analysis for the given threshold (using R).

* Objective → 1. To understand the concept of Association Rule Mining
2. To understand and implement the Apriori algorithm.

* Theory →

Market Basket Analysis is one of the key techniques used by large retailers to uncover associations "btw" items. It works by looking for combinations of items that occur frequently in transactions. Association rules are widely used to analyze retail basket or transactions data, and are intended to identify strong rules discovered in transaction data using measures based on association rules. It is an unsupervised data mining technique to identify items which are bought together. It creates an association if it is for the given scenarios.

• Applications:

1. Music recommendations
2. Medical diagnosis - like diabetes
3. Content optimization - like in magazines, websites or blogs.
4. Trending items that customers buy
5. Customized emails with add-on sales

• Association Rule Mining (ARM):

It is used when you want to find an association "btw" different objects in a set, find frequent patterns in a transaction or database. It is widely used in Marketing, Basket Data Analysis in retailing, clustering and classification.

ARM is viewed as a 2-step approach →

1. Frequent Itemset Generation - Find all frequent item sets with support \geq pre-determined minimum support
2. Rule Generation - List all association rules from frequent itemset. Support and confidence are calculated for the rules.



Let A and B be products, and N be the total products.

$$\text{Support} = \frac{\text{freq}(A)}{N}$$

$$\text{Confidence} = \frac{\text{Support}(A \cup B)}{\text{Support}(A)}$$

$$\text{Support}(A \cup B) = \frac{\text{freq}(A \cup B)}{N}$$

Confidence shows the strength of association.

$$\text{Lift} = \frac{\text{Support}(A \cup B)}{\text{Support}(A) \times \text{Support}(B)}$$

$\text{Lift} = 1 \rightarrow A, B$ are independent; no rules can be derived

$\text{Lift} > 1 \rightarrow A, B$ are dependent

$\text{Lift} < 1 \rightarrow$ More randomness

• Apriori Algorithm:

It is given by R Agrawal and R Srikant in 1994 for finding frequent itemsets in a given dataset, for Boolean Association rule. We apply an iterative approach on level-wise search, where k -frequent itemsets are used to find $k+1$ itemsets.

To improve the efficiency of level-wise generation of frequent itemsets, an important property called the Apriori Property is used. It helps by reducing the search space.

• Apriori Property:

All subsets of a frequent itemset must be frequent. If an itemset is infrequent, then all its supersets will be infrequent.

* Conclusion \rightarrow Thus in this assignment, we implemented the Apriori algorithm to generate association rules; and understood the concept of Association Rule Mining.