

## ABSTRACT

As enormous amount of data is stored in the database and data warehouse, one of the greatest challenge facing individuals and organizations is how to turn their expanding data collection into accessible and actionable knowledge. Data mining is the best-known technique used for it .It is used both to increase revenues and to reduce cost .The machine learning technique perform poorly when applied to the large datasets. Association rules are powerful technique used to show relationships between data items. In this process we discover a set of association rules at multiple level of abstraction from the relevant set(s)of data in the database.

Apriori algorithm is the first best-known algorithm for association rule mining. After some careful examination, we believe that the bottleneck of the Apriori method is at the candidate set generation and the testing that weather an item set is subset of a given transaction. If one can avoid generation of huge set of candidates, and time to find weather a item set is the subset of the given transaction. The mining performance can be substantially improved.

We develop a Signature-based method in which we generate the signature of every transaction and when we have to find weather the given item set is the subset of the given transaction, we simply do the bit wise AND of the signatures of the transaction and the item set. This reduces the time to find the support of the item set, which is usually much less then the subset finding method of apriori method. The number of database scans in the signature algorithm is only one as compare to the database scan for each iteration in case of Apriori algorithm. In the Database pruning method as the database is pruned after each iteration, so each time we have to scan less number of transactions that make it faster than the Apriori algorithm.

Various advantages of the existing algorithms can be coupled to build new algorithms. We also aim to remove repeated scans of irrelevant data in the database by efficient pruning. We have evaluated the performance of the algorithm on the real time data sets provided by “CHADHA PROVISION STORE” F-14/45,MODEL TOWN-II DELHI-10009