IMPLEMENTATION OF MARKET BASKET ANALYSIS USING APRIORI ALGORITHM

V.PARKAVI[1],

II-M.Sc.Computer Science,

Department of computer science,

Dr.N.G.P Arts and science college,

Coimbatore.

Mrs.V.SHOBANA[2,]

Assistant Professor,

Department of computer science,

Dr.N.G.P Arts and science college,

Coimbatore.

Abstract— **Market basket analysis is one of the interested research areas that has received more attention by researchers. Market Basket Analysis also known as affinity analysis or association rule learning, is a data mining technique that can be used in various fields. The main aim of MBA in marketing is to provide the information to the retailer to understand the purchase behaviour of the buyer, which can help the retailer in correct decision making. Mining frequent item sets to build classifiers with good accuracy is essential for efficient algorithm. There are various algorithms are available for performing Market basket analysis. To find the purchase patterns, simultaneously knowing what products are often purchased can be used data mining techniques. This paper discusses the Association rule mining using Apriori algorithm which may helpful to examine the customer behaviour and assists in increasing the sales.**

Keywords— **Datamining, Association rule mining, knowledge discovery and data, Apriori algorithm**.

I. INTRODUCTION

Today, the large amount of data is being maintained in the databases in various fields like retail markets, banking sector, medical field etc. But it is not necessary that the whole information is useful for the user. That is why, it is very important to extract the useful information from large amount of data. This process of extracting useful data is known as data mining or A Knowledge Discovery and Data (KDD) process. The overall process of finding and interpreting patterns from data involves many steps such as selection, preprocessing, transformation, data mining and interpretation. Data mining helps in the business for marketing [2].

The knowledge of the relation between the items can use association rules technique. Type of supervised is included in Association rules technique. The type of Supervised found that the knowledge is based on the generated rule. The system of the relation from each item in the basket can be the knowledge to determine the future strategy obligation. There are so many methods are available to determine the relation of the items by combining any algorithms [1]. The invention of the new system or knowledge can compare some techniques for a datum therefore the new analysis for the invented knowledge is needed.

Predictive analytics is done based on predict & analysis, but it works in reverse viz. first analyze then predict. A retailer must know the needs of customers so that the retailer can adapt the customers. Market basket analysis is a way to find out which items can be purchased together. Market basket analyses gives the retailer good information about items and sales. On group of goods basis Customers who buy bread often also buy several products related to bread like milk, butter or jam. It makes sense that these groups are placed side by side in a retail center so that customers can access them

quickly. Such related groups of goods also must be located side-by-side in order to remind customers of related items and to lead them through the center in a logical manner [3]. This research discusses the market basket analysis by using apriori algorithm.

II. Research Methodology

A. Data Mining

Data mining or often referred to as knowledge discovery in databases (KDD) is an activity that includes gathering, using historical data to find order, patterns or relationships in large data. This data mining output can be used to help make future decisions. The development of KDD has caused the use of pattern recognition to diminish because it has become a part of data mining [10].

B. Association Rule

Association rule mining is a technique to identify frequent patterns and associations among a set of items. For example, understanding customer buying habits. By finding correlations and associations between different items that customers place in their 'shopping basket,' recurring patterns can be derived [5].

Fig: Frequent itemsetmining[8]

$$Support(\{X\} \rightarrow \{Y\}) = \frac{Transactions\ containing\ both\ X\ and\ Y}{Total\ number\ of\ transactions}$$
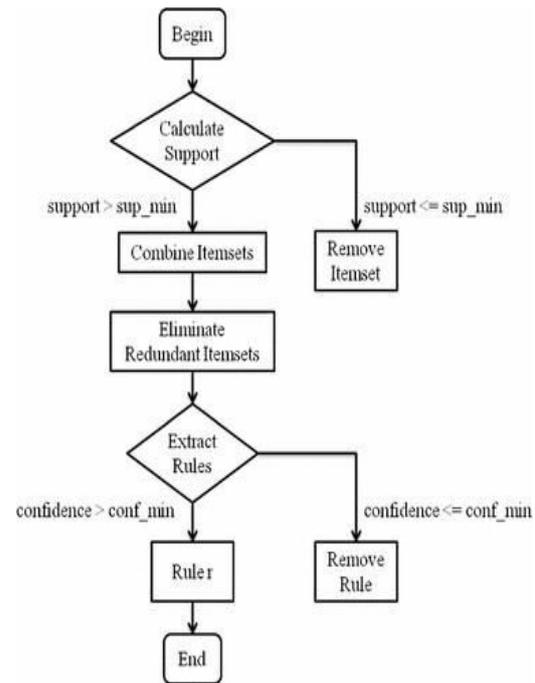
There are three common metrics to measure association mining. They are,

A. *Support:* It is an indication of how frequently the items appear in the data. (i.e) support is the fraction of the total number of transactions in which the item set occurs.

B. *Confidence:* It indicates the number of times the if-then statements are found true. Confidence is the probability of occurrence of consequent.

$$Confidence(\{X\} \rightarrow \{Y\}) = \frac{Transactions\ containing\ both\ X\ and\ Y}{Transactions\ containing\ X}$$

C. *Lift:* It can be used to compare confidence with expected confidence. This says how likely item Y is purchased when item X is purchased, while controlling for how popular

item   Y   is   [4].
Mathematically,

$$Lift((X) \rightarrow (Y)) = \frac{(Transactions\ containing\ both\ X\ and\ Y)/(Transactions\ containing\ X)}{Fraction\ of\ transactions\ containing\ Y}$$

III.APRIORI ALGORITHM :

 A.*Apriori algorithm overview:*

In 1994, Agrawal and Srikant proposed the Apriori algorithm. Apriori is considered to work on databases containing (for example, collections of items bought by customers, or details of a website frequentation) itemsets. Every transaction is considered as a set of items. Given a threshold , the Apriori algorithm identifies the item sets which are subset of at least transactions in the database.

The approach used by Apriori is a "bottom up" approach, where frequent subsets are extended one item at a time, and groups of candidates are tested against the data. The algorithm terminates when no further transactions are found. Apriori uses breadthfirst search and a Hash tree structure to count candidate item sets efficiently [6].

B.*Benefits of apriorialgorithm* :

1. Use large itemset property
2. It is easily paralleized
3. It is easy to implement

C.*Limitations of apriorialgorithm* :

1. It requires many database scans.
2. If the dataset is small, the algorithm can find many false associations that happened simply by chance. It can be addressed by evaluating obtained rules on the held-out test data for support , confident, lift and conviction values.

D.*Working of Apriori Algorithm*

In generation of frequent itemsets we are considering the following implementation scenario on data fetched from web server log file. The Dataset is the transactional data fetched from a Live Web Server Log File. Consider the database taking the support value as 35%. By which every itemset should appear more than 35% in dataset.

Step 1: First we have to Create a frequency table of all the items.

Table1: Frequency table

Item	Frequency
Milk(M)	4
Bread(B)	5
Butter(Bu)	4
Jam(J)	4
Beer(Be)	2

Step 2: We know that the support is greater than or equal to the threshold support. Thus, support threshold is 60%, hence only those items are considerable that occur in further than three transactions and such items are Milk, Bread, Butter and Jam. Therefore, we are left with:

Table2: Frequency Table with Minimum Support

Item	Frequency
Milk(M)	4
Bread(B)	5
Butter(Bu)	4
Jam(J)	4

The table 2 represents the items that are purchased by the consumers frequently.

Step 3: The next step is to make all the possible pairs of the significant items keeping in mind that the order doesn't matter, i.e., AB is same as BA. To do this, take the first item and pair it with all the others such as MB, MBu, MJ. Similarly, consider the second item and pair it with preceding items, i.e., BBu, BJ. Considering the previous itemset as MB (same as BM) already exists. So, all the pairs in our example are. MB, MBu, MJ,BBu,BJ and BuJ.

Step 4: We will now count the occurrences of each pair in all the transactions.

Step 5: Again only those itemsets are significant which cross the support threshold, and those are MB, MBu, BBu, and BJ.

Table 3: Frequency table with 2-Itemset

Item	Frequency
MB	4
Mbu	3
MJ	2
MB	4
BJ	3
BuJ	2

Step 6: Now let's say we would like to look for a set of three items that are purchased together. We will use the itemsets found in step 5 and create a set of 3 items.

Self-join is required for creating a set of 3 items. It tells that from the item pairs MB, MBu, BBu, and BJ . Look for two pairs with the equal first letter and so we get the result.

- MB, MBu , this gives MBBu
- BBu and BJ, this gives BBuJ

Next, we find the frequency for these two transactions.

Itemset	Frequency
MBBu	4
BBuJ	3

Threshold rule is applied again, the only significant itemset that satisfies minimum support Threshold is MBBu. Thus, the most frequently purchased items is MBBui.e Milk, Bread and Butter. As there are no further combinations possible then the Apriori terminates . [11]

IV. DISCUSSION

In a hypothetical study, it is difficult to find frequent factors among the algorithms due to their erratic structural aspects. Hence the distinctive features are taken for discussion besides the advantages and limitations of some algorithms , Apriori algorithm is more efficient during the candidate generation process. It reduces the computation, I/O cost and memory requirement because of the new pruning technique [12]. The visualization of the result is more effortless using this algorithm.

V.CONCLUSIONS

At present many data mining algorithms have been developed and applied on variety of practical problems. In most cases, Associative classification achieves better accuracy compared to some traditional rule-based classification. There are different approaches to rule discovery, pruning, prediction, ranking and evaluation. Apriori algorithm is highly competitive when compared with other traditional classification algorithm.

VI. REFERENCES

[1]. Afrin F Pathan,SnehalPalande, TejalShende,RuchitaPatil, Prof.Vitthal S Gutte Department of Information Technology, MIT college of Engineering, Pune, India ,"A STUDY ON MARKET BASKET ANALYSIS AND ASSOCIATION RULE MINING", Proceedings of National Conference on Machine Learning, 26th March 2019.

[2].ManpreetKaura, Shivani Kanga* aBhaiGurdas Institute of Engineering and Technology, Sangrur 148001, India, "Market Basket Analysis: Identify the changing trends of market data using association rule mining", International Conference on Computational Modeling and Security (CMS 2016) .

[3].LoraineCharlet Annie M.C.1 and Ashok Kumar D2 ,Department of Computer Science, Government Arts College Trichy, India , "Market Basket Analysis for a Supermarket based on Frequent Itemset Mining", IJCSI

International Journal of Computer Science Issues, Vol. 9, Issue 5, No 3, September 2012.

[4].https://blog.usejournal.com/association -rule-mining-apriori-algorithmc517f8d7c54c

[5].https://intellipaat.com/blog/datascience-apriori-algorithm/[6].https://en.wikipedia.org/wiki/Apriori_a lgorithm.

[6].https://en.wikipedia.org/wiki/Apriori_algorith m

[7]. Roshan Gangurde1, Dr.Binod Kumar2, Dr. S. D. Gore3 Research Scholar, Department of Computer Science, SavitribaiPhule Pune University, Pune, Maharashtra, India1 Director, Jayawant Institute of Computer Applications, Pune, Maharashtra, India 2 Rtd. Professor, Department of Statistics, SavitribaiPhulePune University, Pune,Maharashtra, India3, "Building Prediction Model using Market Basket Analysis" , International Journal of Innovative Research in Computer and Communication Engineering,Vol. 5, Issue 2, February 2017.

[8].https://www.google.com/search?q=frequent+itemset +mining&sxsrf=ACYBGNS QO1v12GFxAHuXevBY6jVzpfXhA:1578385048230&

source=lnms&tbm=isch&sa=X&ved=2ahUKEwiAiPnA hvHmAhUUzzgGHXLOCUMQ_AUoAnoECA8QBA& cshid=1578385182025499&biw =1366&bih=657#imgrc=arUshOqmBg2rm M:

[9].https://www.datasciencecentral.com/pr ofiles/blogs/the-7-most-important-datamining-techniques.

[10].Implementation of Apriori Algorithm for Analysis of Consumer Purchase Patterns, The International Conference on Computer Science and Applied Mathematic, Suprianto Panjaitan1, Sulindawaty2 , Muhammad Amin3, Sri Lindawati 4, Ronal Watrianthos5, HengkiTamando Sihotang6, Bosker Sinaga7

[11] DATA MINING APRIORI ALGORITHM IMPLEMENTATION USING R , D Kalpana, International Research Journal of Engineering and Technology (IRJET), Volume: 04 Issue: 11 | Nov -2017

[12] A Survey on Association Rule Mining T. Karthikeyan1 and N. Ravikumar2 ,International Journal of Advanced Research in Computer and Communication Engineering ,Vol. 3, Issue 1, January 2014.