AN OVERVIEW OF MARKET BASKET ANALYSIS USING APRIRORI ALGORITHM

[1]Ms.P.Nithya, M.Sc., M.Phil.,[2]K.Sivapriya, [3]M.Rajshree

[1]*Assistant Professor*[2][3]*M.Sc(IT) Stdents.*

Department of Compute Science & Information Technology

Nadar Saraswathi College of Arts & Science , Theni, India.

Abstract

We live in a world where vast amounts of data are collected daily. Analyzing such data is an important need. data mining can meet this need by providing tools to discover knowledge from data. In addition, many other terms have a similar meaning to data mining—for example, knowledge mining from data, knowledge extraction, data/pattern analysis, data archaeology, and data dredging. Many people treat data mining as a synonym for another popularly used term, knowledge discovery from data, or KDD, while others view data mining as merely an essential step in the process of knowledge discovery. Data mining is the process of discovering interesting patterns and knowledge from large amounts of data. The data sources can include databases, data warehouses, the Web, other information repositories, or data that are streamed into the system dynamically. In this work we mainly focused study on market basket analysis. It is a data mining technique used by retailers to increase sales by better understanding customer purchasing patterns. It involves analyzing large data sets, such as purchase history, to reveal product groupings, as well as products that are likely to be purchased together.

I. INTRODUCTION

Data mining is the process that uses a variety of data analysis tools to discover pattern and relationships in data that may used to make valid prediction. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems[1]. Data mining tools can answer business questions that traditionally were too time consuming to resolve.

Data Mining (DM) is the process of analyzing data from different perspectives and summarizing it into useful information. It is the efficient discovery of valuable, non-obvious information from a large collection of data. It is a knowledge discovery process helps us to understand the substance of the data in

special unsuspected way Data mining consists of extract, transform, and load transaction data onto the data warehouse system, Store and manage the data in a multidimensional database system, Provide data access to business analysts and information technology professionals, Analyze the data by application software, Present the data in a useful

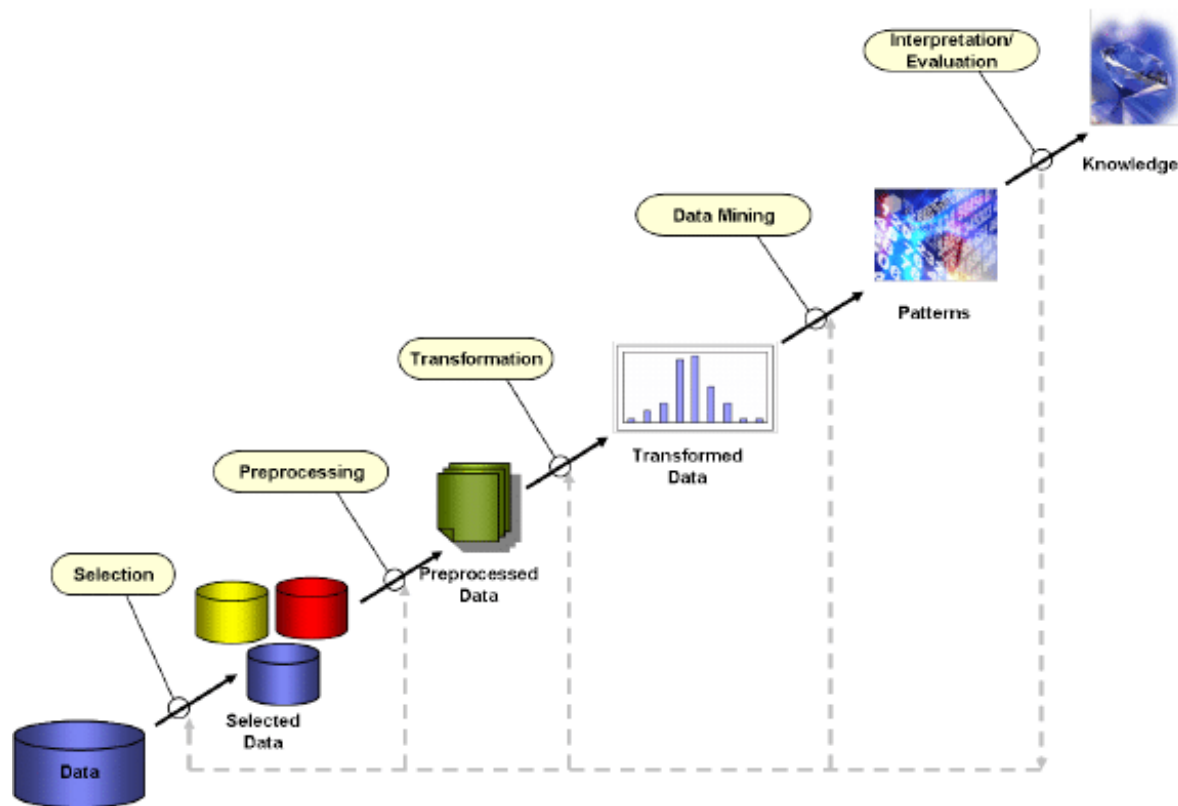format, such as a graph or table. [2]

Fig 1: KDD Process.

Here is the list of steps involved in Knowledge discovery process:

- Data cleaning
- Data Integration
- Data Selection
- Data Transformation
- Data Mining

- Pattern Evaluation

- Knowledge Presentation

II.   MARKET BASKET ANALYSIS

Market Basket Analysis is one of the key techniques used by large retailers to uncover associations between items. It works by looking for combinations of items that occur together frequently in transactions. To put it another way, it allows retailers to identify relationships between the items that people buy[3].

Association Rules are widely used to analyze retail basket or transaction data, and are intended to identify strong rules discovered in transaction data using measures of interestingness, based on the concept of strong rules.

Association Rule Mining

Association rules mining is an important branch of knowledge mining research. They are used for finding frequent patterns and associations among sets of items in transactional databases, relational databases, and other information repositories [4]. An association rule is the relationship between two disjoint itemsets, X and Y. An association rule is of the form: - X -> Y X => Y: - When X occurs, Y also occurs. Given a set of items $I = \{I1, I2,…,Im\}$ and a database of transactions $D = \{t1, t2,......,tn\}$ where $ti = \{Ii1, Ii2…. Iik\}$ and $Iij \in I$, an association rule is an implication of the form X=>Y where $X, Y \subseteq I$ are sets of items called item sets and $X \cap Y = \emptyset$. Association rule mining has been used in a retailing where discovering of purchase patterns between products is very useful for decision making.

Frequent Itemset

Frequent pattern analysis allows a researcher to systematically identify patterns that emerge from database. Frequent pattern mining comprises frequent item set mining and association rule induction[5].

Market Basket Analysis

Market Basket Analysis is a knowledge mining technique that is widely used to identify consumer patterns such that if customer buys certain group of items then customers are likely to buy another group of items[6]. Market basket analysis is an important component in retail organizations. It is a very useful technique for finding out co-occurrence of items in consumer shopping baskets.

Support

It is the measure of how often the collections of items in an association occur together as percentage of all transactions. Support(s) for an association rule $X => Y$ is the percentage of transactions in the database that contains X U Y. Every association rule has support. The rule that has very low support may occur simply by chance.

Confidence

Confidence for an association rule X=>Y is the ratio of the number of transaction that contain both antecedent and consequent to the number of transaction that contain only antecedent. A rule with low confidence is not meaningful. Confidence (α) for an association rule x=>Y is the ratio of number of transactions that contains X U Y to the number of transactions that contains X.

Minimum Threshold Values

The strength of an association rule can be measured in terms of its support and confidence. The rules derived from itemsets with high support and high confidence. The number of association rules that can be derived from a dataset are large. Interesting association rules are those whose support and confidence are greater than minimum support and minimum confidence[7]. The number of association rules discovered is affected by a user's decision concerning the minimum support threshold and minimum confidence threshold. support and confidence values occur between 0% and 100%.

III.   **DATA MINING TECHNIQUE USED FOR MARKET BASKET ANALYSIS**

Apriori algorithm for finding frequency item set

The Apriori algorithm analyses a data set to determine which combinations of items occur together frequently. It is at the core of various algorithms for data mining problems. The best known problem is finding the association rules that hold in a basket-item relation. Basic idea behind this algorithm is

- An item set can only be a large item set if all its subsets are large item sets.
- The sets of items that have minimum support can be considered.
- Association rules can be generated from frequent item sets[8].

Table 1 describe sample data set of Customer,

TID	ITEM1	ITEM2	ITEM3
TXl	Bread	Butter	Milk
TX2	Ice-cream	Bread	Butter
TX3	Bread	Butter	Noodles
TX4	Bread	Noodles	Ice-cream
TX5	Butter	Milk	Bread
TX6	Bread	Noodles	Ice-cream
TX7	Milk	Butter	Bread
TX8	Ice-cream	Milk	Bread
TX9	Butter	Milk	Noodles
TX10	Noodles	Butter	Ice-cream

Teble 1: Sample Data Set.

Step 1: First Step is to find support of each items on dataset.

Items	Support
Bread	8
Butter	7
Ice-Cream	5
Milk	5
Noodles	5

Table 2:Support

Step 2: Second step is to fix a max support for frequent items which is find in data set and eliminate less supporting set(Table 3)

Items	Support
Bread	8
Butter	7
Ice-Cream	5
Milk	5
Noodles	5

Table 3 :Eliminating less support item.

Step 3 : Find a frequent item set for two items in data set.(Table 4)

Item-sets	Support
{Bread, Butter}	5
{Bread,Ice-cream}	4
{Bread, Milk}	4
{Bread, Noodles}	2
{Butter, ice-cream}	3
{Butter, Milk}	4
{Butter, Noodles}	3
{Ice-cream, Milk}	1
{Ice-cream,Noodls}	3
{Milk, Noodles }	3

Table 4 : Frequent Item set.

Step 4: Eliminate minimum support frequent itemset which are not found regularly .(Table 5)

Item-sets	Support
{Bread, Butter}	5
{Bread, Milk}	4
{Bread, Noodles}	2
{Bread, ice-cream}	3
{Butter, Milk}	4
{Butter, Noodles}	3
{Noodles, ice-cream }	3

Table 5: Pruning frequent itemset

Step5: Find a frequent item set for three items in data set.(Table 6)

item-sets	support
{Bread, Butter, Milk}	3
{Bread, Butter, Noodles}	1
{Bread, Milk, ice-cream}	1
{Bread, Butter, ice-cream}	0
{Butter, Milk, Noodles}	1
{Bread, Milk, Noodles}	0
{Noodles, ice-cream, Bread}	2

Table 6: Frequent Itemset

Step 6: Pruning frequent itemset which are no find frequently (Table 7)

Item-sets	Support
{Bread, Butter, Milk}	3

Table 7: Pruning Frequent Itemset

IV.   CONCLUSION

Market basket analysis is widely used for predicting frequent items .the data set for market basket analysis has set of customer transactional data .In this paper we mainly focused on Apriori algorithm how it was predicting the frequent itemset. Apriori produce association rules for better prediction of customer behavior.

V.   REFERENCE

[1] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," Proc.   20th   Int'l   Conf. Very Large Data Bases (VLDB '94), pp. 487-499, 1994.

[2] http://www.ijarcsse.com/docs/papers/Volume_3/3_March2013/V3I3-0162.pdf

[3] https://datascienceplus.com/a-gentle-introduction-on-market-basket-analysis%E2%80%8A-%E2%80%8Aassociation-rules/

[4] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," Proc.   20th   Int'l Conf. Very Large Data Bases (VLDB '94), pp. 487-499, 1994.

[5] Arpan Shah Pratik A. Patel, "A Collaborative Approach of Frequent Item Set   Mining: A   Survey," International Journal of Computer Applications (0975 –   8887)

[6]Loraine Charlet Annie M.C and Ashok Kumar D, ," 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 ISSN (Online):   1694-   0814.

[7]Anand H.S. and Vinodchandra S.S.," Applying Correlation Threshold on Apriori Algorithm," 2013 IEEE International Conference on Emerging Trends in Computing,   Communication and Nanotechnology

[8] Market basket analysis using apriori algorithm in data mining SWATI MAHESH JOSHI International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 05 Issue: 04 | Apr-2018