A STUDY OF CUSTOMERS' BEHAVIOUR ANALYSIS USING DATA MINING TECHNIQUES

B.Gobinath[1] , Dr.T.K.Shanmugam[2]

[1]Research Scholar, Gobi Arts & Science College, Gobichettipalayam, Tamilnadu, India.
[2]Associate Professor in Computer Science,
Gobi Arts & Science College, Gobichettipalayam, Tamilnadu, India.

ABSTRACT:

Data mining is the field of analyze about Customer Buying habits to research our day to day problems and provide solutions on it. Consumer behavior analysis is an important tool to understand the customers. By looking into consumer psychology and the forces behind customer buying behavior, organization can craft new products, marketing campaigns and increase profitability.   Data mining is the best way to analyze customer activities and their buying habit while purchasing. Data mining is an efficient tool for decision making and analysis of their behavior. There is a necessary thing to choose a proper Mining technique to discover the buying habit of the Customer and mine from the Database or Warehouse in fast and furious manner. The Primary goal of this paper is to analyze the Customer needs, and their behavior, their needs arising in customer's mind and the importance of Data mining to replace the traditional method. It also gives the study of Customer behavior analysis in the Sales Industry.

Keywords: Data mining, Customer, Buying habit, Database, Warehouse

1. INTRODUCTION:

The study of Customers to purchase some items or products from the Supermarkets with the question of the basic necessity to purchase that particular item and the reason to purchase it. It is quite interesting way to analyze the psychological thinking of the customers and their mindset and also analyze the behavior of their buying habits. This will turn up into consolidating the behavior into a Mathematical or Algorithmic approach for finding their heuristic interest in Purchasing. The mechanism of finding to choose a particular item will help the Firms or Supermarkets to increase their strategy of selling the products what the customers liked the most and what the customers avoid to purchase?. This analysis can be done based on several aspects. The basic aspects from the Customers side may be the mentality of the Customers and their buying habits, reasons to choose the particular Product and how frequently the customer chooses the particular or selective products. After finding the buying habit of the Customer, the Organization or firm will give the benefits or their marketing strategies towards the customer to interact them to the firm. This study also deals with the customers' decision making and their marketing strategies interacted with the particular item sets and organization. It also identifies the influence by the organization of the customer.

The Primary goal of finding the Customers interest on choosing the items and its benefits along with the factors like taste, price, quality and quantity. It also provides the facility of giving offers from the organization or firm for the customers' interaction. Customers Attitude is the study of finding their likes based on the Society implications, age groups of the individual customer or peer group of customers based on seasonal changes. Consider an example of an IT employee, they would like to purchase some Formal Cloths, for College students they would like to choose Casuals or Jean Cloths. Peoples in certain regional group may choose traditional

outfits. This will help the manufacturing unit of the firm to produce more quantity of the particular product. This will helps the organization to get high profit.

1.1 Application of Customers Behaviour:

The study of customers habit is to increase the marketing strategy. It may help their productivity. Strategy and skills to develop their business more and more. Some customers are influenced by Television Advertisements, Messages from Social Media, reference from their friends and Trial based methods. After finding this Customers habit, the Organization can post their Advertisements in the Prime time. For Example, the food manufacturing firm may broadcast the Advertisement in Lunch or Evening. If the Textile shop may broadcast their advertisement in the seasonal festivals or in the mid of Television Serials or NEWS. So that, they can catch many customers and customers are in tuned to purchase them.
In terms of Social Marketing Groups catches the groups of peoples who search or browse on Internet.

After finding the behavior of the Customer, we can find the mutual relationship of Organization and Customer and the Customers intelligence, the shop keeper can find the frequent likely items as in one basket. This will really helpful for the customers to choose their item sets and it may reduce time to choose items. Consider an Example, If a Customer likes to purchase Bread frequently, the shop keeper may create a basket which contains Bread, Jam, Milk, Sugar and Coffee Powder. This basket is not only for the particular customer. This basket may like majority of the customers in the society to purchase the Group of Item sets in a single basket. This method of Analysis is called as Association Rule Mining. In the field of Marketing, it is called as Market Basket Analysis (MBA). The well known Market Basket Analysis algorithm is Apriori Algorithm. This algorithm is used to find frequent item sets and increasing the customer buying habits.
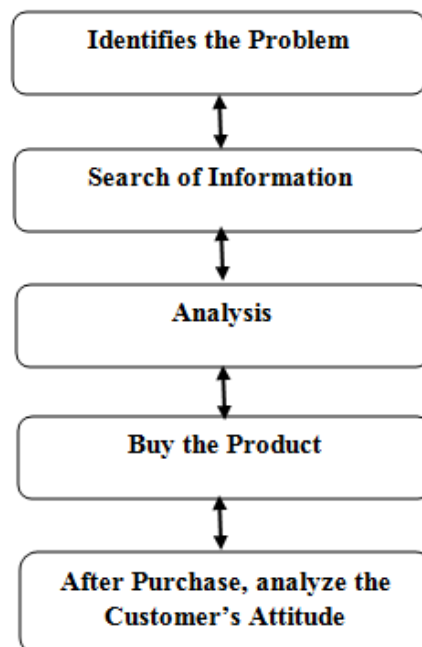
```
┌─────────────────────────┐
│  Identifies the Problem │
└─────────────────────────┘
            ↕
┌─────────────────────────┐
│  Search of Information  │
└─────────────────────────┘
            ↕
┌─────────────────────────┐
│        Analysis         │
└─────────────────────────┘
            ↕
┌─────────────────────────┐
│     Buy the Product     │
└─────────────────────────┘
            ↕
┌─────────────────────────┐
│ After Purchase, analyze │
│  the Customer's Attitude│
└─────────────────────────┘
```

Diagram: Levels of analyzing Customer's behavior

Apriori algorithm is based on collecting the primary data and likes of the customers from multiple source environment. For Example, if a customer purchase a jewel from a well-known jeweler shop at one city. The

customer may also purchase the same jeweler from other city. Apriori algorithm finds the customer anywhere in the country. There are two pitfalls of using Apriori Algorithm.

(i) It analyses the Secular pattern of the Customer
(ii) Lot of Association subsets for every customer.

To overcome these pitfalls, we can move on to Apriori like Algorithm. It is frequently extracts the Association Rules in Multi Platform Environment.

There are lots of views about Customer buying habits using various Data mining techniques. Data mining tools have been the best techniques to research the habits of customers. Some of the problems were occurred while using some data mining techniques. We have to choose a proper algorithm to extract the information from the database or data warehouse. There are various methods like Classification, Cluster analysis and more mining techniques available in Data mining.

There are two levels of mining process in Association Rule Mining algorithm. One is Support Level and another one is Confidence Level. Patterns or item sets should meet the Origin Support level should be there and it should meet the respective Confidence value. The best part of Association rule mining is based on Profit and Quantity. Association rule mining is the field of researching the findings of Facts that may be called as Knowledge Discovery in Database (KDD). Advanced Market based algorithms are used to meet the quality and some of the improved criteria's like quality of sold items and queries from the customers.

Extract the data using Generic Association Rule Mining deals the process of identifying the rules from large data sets using strong Support and Confidence Origin value. Predictive techniques are the key indicators to create some interesting Association rules. This Association rule mining is purely based on developing the Profit of the organization and Business intelligence and identifies the future opportunities of the organization. It also identifies the trending Marketing Strategy and insist the shop keeper to do some important business decisions. It provides Security, benefit, Cost and item sets. There are two stages of using Association Rule Mining. Those are

(i) Finding frequent item sets
(ii) Generate a Proper Association Rules

The best association rule mining for customer behavior analysis is based on learning and it replaces some unstructured decision makings and measuring manual results. It provides the mathematical or algorithmic based approach to apply tools and algorithms for the Structured Decision making.

2. METHOD FOR IDENTIFYING CUSTOMER BUYING HABIT:

Data mining and Warehousing is most essential in various fields in the world. Data mining technique is the mechanism of researching the Customer Buying Habits and increasing the buying habit of the Customers. Analysis and Research of Customers are much important in all the fields and handle billions of active customers and maintain their interests in a database. Collected data from the primary source and keep the data at the backend where the customer does not know the analysis of his/her data and provide some interesting offers especially for the particular customer on their Birthday/anniversary and special day offers. The customer was surprised and the customer will maintain a strong bonding with the organization.

The Definition of Data Mining stands for Knowledge Discovery in Database (KDD). Knowledge Discovery is the technique for finding relevant, useful and interesting patterns from the large repository of Database. The view of Data mining is to collect the needful data that is available in the database or data warehouse. Data mining is the method of finding huge amount of data from the huge inter-related data warehouse. Fundamentally

Data mining is applicable in many fields like buying and selling, banking and finance and networking industries. Analysis of Customer behaviour intends to purchase a product along with another.

2.1 Classification Techniques:

Classification is the mechanism of dividing the item sets based on its characteristics. It is a large group of data sets that is having item sets. There are lots of techniques like Clustering, Detection of Outlier analysis and Evaluation. Generally mining techniques are based on the facts, data abstraction, and knowledge. Data mining system is also generalizing that mining of data based on generally received patterns and exceptions or outlier. Generally, Logic, Association mining, Classification rules, Clustering and Outlier analysis will easily identify the detect analysis and outlier errors.

Data mining technique could be divided based on the methods employed in the application. This approach may be explained based on level of Customer response and their interaction and the usage of database involved in the system like Relational, Spatial, temporal, Relational, Object oriented, Object based or Data warehouse. There are many presentation methods like Mathematical modeling, Chart or Graph based methods and machine learning. The Data mining system can also enable various Data mining techniques that incorporates in a single application to execute efficiency in cost, effort, time and integrated . Data mining techniques can be easily adopted in all fields like banking, finance, tele-communication, education and medicine.

2.2 Definition of Item set:

A Collection of items that are combined together is called as item set. It consists of minimum of two or more items. For Example, an item, rice which can be combined with Dhal. {Rice and Dhal} is called as item set.

2.3 Definition of Frequent Item set:

A set of items which will occur frequently that must satisfy the minimum support threshold value of Support and Confidence level is called as frequent item set. The frequent item set can be combined with any of the item set. For example {Rice, Dhal} can be frequently occurring with Vegetable item set.

3. ACTUAL IMPLEMENTATION:

There are lots of techniques in Data mining to undergo the research about the attitude of the customers and analyze the likeliness of the customer that will help the organization to grow up. Data mining tool collects all the data concurrently about each customer over some methodology and gives the transaction of the customer in day wise, week wise and monthly wise. Database administrator selects the proper method, algorithm and the repository of customer transaction. Administrator can form the proper Association rule mining for providing tangent response between customer and organization.

3.1 Association Rule mining:

Association rule mining analyze and give the relation among huge collection of data items and the analysis will become quite interesting to form the rules and mining technique from the respective company's database . This finding may gives more interestingness among the association relationship in high set of user's transaction data will give an exact solution to make appropriate business decisions like brochure making, attractive card design, tele-marketing and offers towards customers.

The best approach of Association Rule Mining is Market Basket Analysis. This approach is to increase and enhance the customers attitude towards purchase. It also finds the interestingness about customer and gives the

solution to choose the likelihood of products in their own basket. Consider an example of Amazon and Flipcart, it offers variety of Products with various categories like Groceries, Furniture, Cloths ..etc. Users can view the products and purchase. If user feels to purchase later he/she can add the particular item in the cart for later purchase. Likewise the strategy implemented in business transactions we can form the basket which may have the combination of related items like bread with jam and butter or milk with sugar and coffee powder. Basket items should be inter-related because every product in the basket is mandatory to make the cooking. These combinations are liked by major groups in the society.

It may save the customer Time and Effort. This method can intend to improve the marketing strategy can be extended to purchase the related items in a space can make the customers with spending quality time in their transaction. We can form the basket with two important criteria's. First thing is the particular basket should meet all type of peoples in the society. It should have high support level Second thing is the Confidence level. It meets the customer satisfaction and give more confidence to the users to purchase the basket.

Consider the classical example of Apriori Algorithm for a Supermarket application. In a supermarket, there are lots of Products showcased for Selling. The Shopkeeper analyses the Customers buying habit and make the basket which contains the items in a related manner. Administrators have to calculate the Support level and Confidence level for the Item sets in a basket. Support level deals with the value based on the customers need and frequently purchased for the frequent item sets. Confidence level deals with the at most confidence from the customers perspective. In association rule mining, there is a formula to find support level and confidence level of the given item sets will be the minimum support value.

Apriori Algorithm gives three conditions. Probability of any particular item 'I' is frequent where if

(i) $P[I]$ is lesser than the minimum support threshold
(ii) $P[I]+A$ is lesser than minimum support threshold value but A is not a frequent itemset.
(iii) If the item set and its value is lesser than minimum support value, then all the superset will fall under minimum support value.

There are two fundamental steps to be followed in Apriori algorithm. One is Join step and another one is Prune step. Join step is used to generate k+1 item set from k-item set by combining every item with itself. Prune step is used to analyze the count of every items in the database.

The formula for finding support level is:

$$\text{Support level of A, B as Probability}(A \cup B) = \frac{N(A \cup B)}{N}$$

Notaations represented as: **A and B are item sets**
N is the number of items

Here, we can form the item sets available in the super market. A and B are represented as the Items in the Supermarket. N is represented as the number of items in the shop. Probability of (A U B) is the statistical approach to find the frequent patterns of A and B is available in the item sets. For example A={M,N,O} and B={S,T} and the $P(A \cup B) = \{(M,S), (N,S), (O,S), (M,T), (N,T), (O,T)\}$. The value of $P(A \cup B) = 6$.

Likewise the formula of confidence level is:

$$\text{Confidence of A and B as Probability(A/B)} = \frac{N(A \cup B)}{N(A)}$$

Notations represented as Probability of (A/B) as A tense to B
N is represented as number of items

3.2 Frequent Pattern Mining:

Frequent Pattern Mining algorithm is an essential methodology in Data mining technique to find the relationship between various items in the dataset. The Pattern mining relation will represent the form of Association rules. It can easily find the irresponsible item sets in the supermarket. Frequent Pattern mining is applied in many of the applications like data science, software industries, tele-marketing and marketing strategies.

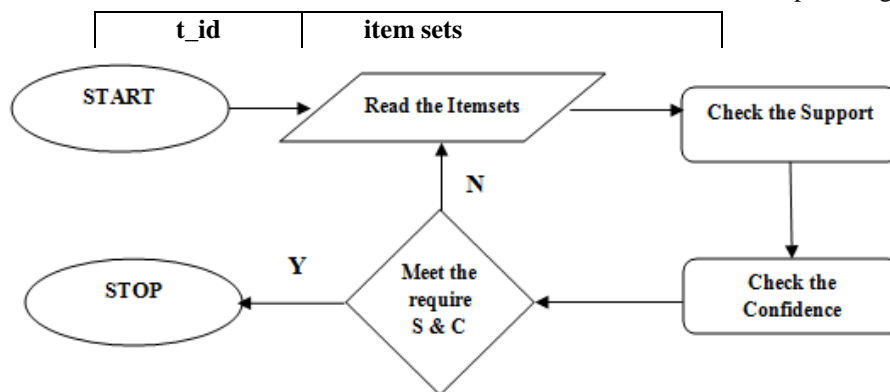The work flow   of Apriori algorithm is:

Diagram: Work flow of Apriori Algorithm

3.3 Association Rules:

There are some pre-defined rules in Association rule mining. That is used to find the relationship between attributes and huge databases. Let we define the basic association rule A=>B will be the form of "In a set of transactions, some values of item set A determines the values of item set B under the strict support and confidence level should be there".

The representation of support and confidence level of the particular item set {Rice, Dhal} as

Rice=>Dhal [Support Level=45%, Confidence Level=60%]

4. EXAMPLE OF IMPLEMENTING APRIORI ALGORITHM:

Consider the example in a supermarket consist of various items and transactions by the customers by analyzing the behavior of their frequent transactions. Consider the table with six transactions and every transactions are made with a transaction_id

T1	{Dhal,Oil,Sugar}
T2	{Sugar,Salt}
T3	{Rice,Dhal,Sugar}
T4	{Rice,Dhal,Oil,Salt}
T5	{Rice,Dhal,Oil,Sugar}

The problem itself, we can choose the minimum threshold value of Support level will be 50% and Confidence level will be greater than or equal to 60%.

The formula for calculating minimum support threshold value is:

$$\text{calculating Support threshold} = \left[\frac{\text{Min Support threshold}}{100} \right] X \text{ No of items}$$

So that: [50/100] X 6 will be 3, we can come to decision that the minimum support value 3 and below will be eliminated for further transactions. Because those item sets may not be liked most by the customers.

Step 1:   Count of each item occurred in the
transaction   table:

t_id	item sets
Item	**No of times occurred**
Rice	4
Dhal	5
Oil	4
Sugar	4
Salt	2

Step 2:

In the Pruning step we can eliminate the item Salt which does not satisfy the minimum threshold value Salt<=3. Now the transaction table contains four items. {Rice,Dhal,Oil,Sugar}

Step 3:

Item sets	No of times occurred
Rice,Dhal	4
Rice,Oil	3
Rice,Sugar	2
Dhal,Oil	4
Dhal,Sugar	3
Oil,Sugar	2

Now calculate the possibilities of items combinations and find the count of occurance in the transaction table.

Repeat Step 2. Now we prune the table which the occurrence is lesser than or equal to 3. {Rice,Sugar} and {Oil,Sugar} gets eliminated. Because these item sets had not liked by most of the customers.

Step 4:

Now we conclude the Items in all possible of combinations: {Rice,Dhal,Oil}, {Rice,Dhal,Sugar}, {Rice,Oil,Sugar} and {Dhal,Oil,Sugar}.

Step 5:

Now, generate Association rules and discover all possible combinations of four item sets.

{Rice,Dhal} => {Oil}, Confidence level=[Support(Rice,Dhal,Oil)/Support(Rice,Dhal)] * 100

Confidence(Rice,Dhal,Oil) = (3/4)*100 => 75%

{Rice,Oil} => {Dhal}, Confidence level=[Support(Rice,Oi,Dhal)/Support(Rice,Oil)] * 100

Confidence(Rice,Oil,Dhal) = (3/3)*100 => 100%

Likewise, we calculate the Confidence level and finally eliminate the Confidence value is lesser than or equal to 60%.

In final, we conclude the high Confidence value is the best combination that most of the customers like the transaction item sets.

5. ADVANTAGES AND DISADVANTAGES OF ASSOCIATION RULE MINING:

It has various advantages as well as disadvantages. Advantages are Association Rule Mining is easy to learn and implement. It has only two steps that are Join and Prune. The disadvantages are very complex while using larger transactions, Another disadvantage is every time the entire transaction table should be pruned.

6. CONCLUSION:

Consumers' behaviour has placed an important role over 10 years. It is becoming a special criterion of unique quality, access to knowledge, which is a measurable source of progress. Consumer behavior is an important as far as buyers' behavior is concerned, and it is attracting growing interest from researchers. It is used to provide

great flexibility and susceptibility to changes. Association Rule Mining is the efficient technique in commercial based applications. It reduces the size of the item sets based on minimum threshold value. The efficiency is to scan every time, we can get the accurate support by the customers. In this method, we can easily find the heuristic likeliness of the customers. This paper analyses on consumer buying decision-making process based on the analysis. The customer purchase goods or services with low involvement in decision making if the product is routine usage. For buying daily usage products, the most probability is to skip the one or more than one stages in the decision making process.

REFERENCES:

[1] Jagmeet Kaur, Neena Madan, "Association Rule Mining – A Survey" , International Journal of Hybrid Information Technology, 2015.

[2] J.Han,M.Kamber, J.Pei, "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers, 2001.

[3] J.K.Jain, N.Tiwari, M.Ramaiya, "A Survey on Association Rule Mining", International Journal of Engineering, Research, and Applications, 2013.

[4] Kaveh Peighambari, Setayesh Sattari, Arash Kordestani, "Consumer Behaviour Research", SAGE Research, 2016.

[5] N.Padhy, P.Misra,, R.Panigrahi, "The Survey of Data Mining Applications and Future Scope", International Journal of Computer Science Engineering and Information Technology (IJCSEIT), 2012.

[6] Pang Ning Tan, Michael Steinbach, Vipin Kumar, "Introduction to Data Mining", Pearson Education India, 2016.

[7] S.Kotsiantis,D.Kanellopolous, "Association Rules Mining – A Recent Overview", GESIS International Transaction, 2006.

[8] T.Karthikeyan, N.Ravikumar, "A Survey on Association Rule Mining", International Journal of Advanced Research in Computer and Communication Engineering, 2014.

[9] Agrawal R., Imielinski T and Swami A, "Mining association rules between sets of items in large databases", ACM SIGMOD Conference, Washington DC,USA.

[10] Han J, Kamber M and Pei J, " Data mining: Concepts and Techniques" , San Francisco, Morgan Kaufmann Inc (3rd Edition)

Books:

[1] G.K.Gupta, Introduction to Data Mining with Case Studies, PHI Learning Private Limited, Third Edition, 2014.