



# CUSTOMER BEHAVIOUR ANALYSIS USING DATA MINING

Shreyash S. Mane<sup>1</sup>, Shweta M. Barhate<sup>2</sup>, Satish J. Sharma<sup>3</sup>

<sup>1</sup>Student, <sup>2</sup>Assistant Professor, <sup>3</sup>Professor

<sup>1,2,3</sup>Department of electronics and Computer Science,

<sup>1,2,3</sup>RTM Nagpur University, Nagpur, India.

**Abstract:** In India, there are number of entrepreneurs or retailers, who have started their own business or launched their own products. The major difficulty that they were facing is marketing. They spend a lot of money on marketing but do not get an appropriate result as they are not aware about their probable customer, details of the customer, preferability about their products, etc. So, they couldn't achieve much success be it increase in sales, customer satisfaction, etc. Study of customer behavior in retail stores usually deals with identification of customers and their buying behavior patterns. All the activities of the business concerns end with customer attraction and customer satisfaction. Customer behaviour analysis has become an integral part of strategic market planning. To perform the customer behaviour analysis we used an association rule mining and clustering. Association rule mining identifies relationship between a large set of data items. Using association rule mining associativity between two products is made effectively so that the associated item is bought along with the most frequent bought products. Clustering is a technique which forms a group of similar data. Using Clustering technique, we get the cluster of most purchased product. This will help the entrepreneurs, retailers for making marketing strategies and increase product sales to achieve high levels of customer satisfaction. The study was carried out with the help of data mining techniques on retail\_dataset.csv. The experimental work is done in weka software.

**Keywords:** Data Mining, Association rule mining, Apriori algorithm, Clustering, Simple K-Means algorithm.

## I. INTRODUCTION

Marketing is evolving day by day. One can buy lot of products such as stationery, groceries, and electrical appliances. These products can be purchased in stores and on e-commerce websites. A customer is someone who buys a product according to their needs and choices. All marketing activities are related to customer acquisition and customer satisfaction. But the customer attraction and customer satisfaction is a very big problem to sellers. Customer attraction and customer satisfaction is a technique to develop marketing strategies. Seller tried many ways to attract the customers and satisfied them. A seller gives an offer to customers or they do promotions to attract the customers. But still couldn't get the profit as many sellers can, because a seller has lack of knowledge of the customers purchasing behaviour.

Customer purchasing behaviour has become an integral part of strategic market planning. It involves investigating individuals and organizations, and how they choose and use products and services. With the help of customer behavioural analytics, sellers can maintain their stores the way customers want. Which intern helps the seller to know the product that the customer likes most or frequently bought item? The seller can arrange the product and give an exciting offer or discount on the product in order to attract and satisfy the customer. A proven method for analyzing customer behavior is data mining.

The process of extracting available data from a larger set of any raw data is called Data mining or also known as "knowledge discovery in a database". Data mining allows you to understand what is relevant and use that information to get the exact results you expect. There are different techniques for solving different types of problems. A very useful technique to the find the associations and relationships among the large set of data items is association rule in data mining. Data mining is widely use in retail, banking, manufacturing, education, insurance and telecom and media technology.

## II. REVIEW OF LITERATURE

Shashi Pal Singh, et al. [1] uses data mining systems in their paper, to provide an easy way to understand and analyze customer behavior in retail department stores. Their motivation was to analyze product sales patterns according to market demand and increase business profits.

Milan Patel, et al. [2] is focus on key issues for improving decision making. Customer relationship optimization in a highly customer-centric business. This paper showed that data mining can be integrated into customer relationship management (CRM) to improve the CRM process.

Abhijit Raorane, et al. [3] examines consumer buying behaviour when faced with real-world problems. The main focus is on the knowledge of consumer behavior and the psychological state of the consumer at the time of purchase. They used mining association rules to find interesting correlations between large amounts of data items.

M. Alex Syaekhoni, et al. [4] analyzes eco-friendly products with environmental considerations in mind. This survey is based on an analysis of customer purchasing behavior as an alternative to support decisions to promote green products in the retail industry. This model is based on RFID (Radio Frequency Identification). This helps retailers increase product sales and achieve high customer satisfaction.

Muqaddas Gull, et al. [5] focus on analyzing the behavior of people who visit online shopping sites and browse different things. They applied apriori algorithm and association rule of mining for the customer behavior analysis.

M. N. Saroja, et al. [6] analyzed purchases by grouping purchases by gender. They analyzed which group of people would buy more products. By considering this factor, retailers can improve product sales. According to the group, retailers can anticipate the products customers buy most and provide marketing strategies to improve sales.

Savi Gupta et al. [7] use data mining methods to identify relationships between large datasets. It makes sense to look at a customer's buying behavior. This paper describes existing data mining algorithms for shopping cart analysis.

### III. Existing Methodologies

#### 1. Clustering

Clustering is a technique for forming groups of similar data. The purpose of clustering is to find a structure (unique grouping) from a collection of unlabelled data. For analysis, we select the best one between K-means clustering and EM Clustering.

##### A. K-Means Clustering

1. The cluster representation of k-means algorithm is based on the mean.
2. To initialize the cluster, select randomly k-means.

##### B. EM Clustering

1. The cluster representation of EM Clustering algorithm is based on the mean, variance and weight.
2. To initialize the cluster EM Clustering used K Gaussian distributions.

#### 2. Association rule mining

Association rule mining identifies relationship between a large set of data items. For the analysis, we select the best one between Apriori algorithm and Fp-Growth algorithm.

**A. Apriori Algorithm:** - Apriori generates the frequent patterns by making the itemsets using pairing such as single item set, double itemset, and triple itemset.

**B. Fp-Growth Algorithm:** - FP Growth generates an FP-Tree for making frequent patterns.

### IV. EXPERIMENTAL WORK

#### Dataset

	A	B	C	D	E	F	G
1	0	1	2	3	4	5	6
2	Bread	Wine	Eggs	Meat	Cheese	Pencil	Diaper
3	Bread	Cheese	Meat	Diaper	Wine	Milk	Pencil
4	Cheese	Meat	Eggs	Milk	Wine		
5	Cheese	Meat	Eggs	Milk	Wine		
6	Meat	Pencil	Wine				
7	Eggs	Bread	Wine	Pencil	Milk	Diaper	Bagel
8	Wine	Pencil	Eggs	Cheese			
9	Bagel	Bread	Milk	Pencil	Diaper		
10	Bread	Diaper	Cheese	Milk	Wine	Eggs	
11	Bagel	Wine	Diaper	Meat	Pencil	Eggs	Cheese
12	Cheese	Meat	Eggs	Milk	Wine		

Fig. 1: dataset

**Algorithm:**

Step 1. Open WEKA.

Step 2. In the Preprocess tab, click on Open file button and select the dataset.

Step 3. Click on the Associate tab, then click on Choose button in Associator and select Apriori.

Step 4. Inorder to change the parameters for the run (for example support, confidence, etc.) we click on the textbox immediately right of the choose button.

Step 5. Click to start

**Apriori Algorithm**

1. Set the minimum support threshold.
2. Consider every single item as an itemset and count the support from their frequency.
3. Capture only those itemset who have support is equal or greater than minimum support threshold.
4. After getting all the larger itemset, generate another itemset having those itemset
5. Repeat.
6. Until no change.

**Best rules found:**

1. **0 = Cheese 3 = milk 32 => 2 = Eggs 32**
2. **0 = Cheese 2 = Eggs 32 => 3 = Milk 32**
3. **2 = eggs 3 = milk 35 => 0 = cheese 32**

**Algorithm:**

Step 1. Open WEKA.

Step 2. In the Preprocess tab, click on Open file button and select the dataset.

Step 3. Click on the Cluster tab, then click on Choose button in Clusterer and select SimpleKMeans.

Step 4. Inorder to change the parameters for the run (for example number of cluster, distance function, etc.) we click on the textbox immediately right of the choose button.

Step 5. Click to start

**K-Means Clustering**

**Input:** The number of k and a database containing n objects

**Output:** A set of k- clusters that minimizes the squared-error criterion.

**Method:**

1. Arbitrarily choose k objects as the initial cluster centres
2. Repeat
3. (re) assign each object to the cluster to which the object is most similar based on the mean value of the object in the cluster
4. Update the cluster mean, i.e. calculate the mean value of the object for each cluster
5. Until no change

**K-means** =====

Cluster 0: Bread, Wine, Diaper, Eggs, Pencil, Pencil, Bread

Cluster 1: Meat, Diaper, Eggs, Cheese, Bagel, Pencil, Bread

**Clustered Instances**

0 181 (57%)

1 134 (43%)

**V. RESULTS****Apriori Algorithm Result**

**2 = Eggs 3 = Milk 35 ==> 0 = Cheese 32**

Support (Cheese) =  $56 / 315 = 0.177777778 = 0.1$

Confidence =  $0.101587302 / 0.111111111 = 0.914285719 = 0.91$

Lift =  $0.914285719 / 0.177777778 = 5.14285713 = 5.14$

Leverage =  $0.101587302 - (0.111111111 * 0.177777778) = 0.0818342156 = 0.08$

**K-Means Algorithm Result**

In clustering, it groups a product of most purchased product. The clusters are formed with associations between products. It gives us buying tendency of the product.

## VI. CONCLUSION

Association rule mining identifies relationship between a large set of data items. Using association rule mining associativity between two products is made effectively so that the associated item are bought along with the most frequent bought products. Clustering is a technique for forming groups of similar data. Using Clustering technique, we get the cluster of most purchased product. The result of this study is to help the entrepreneurs and retailers for making marketing strategies. It can help in order to increase product sales and achieve high levels of customer satisfaction.

## REFERENCES

- [1]. Shashi Pal Singh, Ajai Kumar, Neetu Yadav and Rachana Awasthi, “**Data Mining: Consumer Behavior Analysis**”, 3<sup>rd</sup> IEEE International Conference on (RTEICT-2018), MAY 18<sup>th</sup> & 19<sup>th</sup> 2018.
- [2]. Milan Patel, Srushti Karvekar and Zeal Mehta, “**Customer Behavior Model Using Data Mining**”, International Journal of Advanced Technology in Engineering and Science Volume No.02, Special Issue No. 01, September 2014.
- [3]. Abhijit Raorane and R. V. Kulkarni, “**Data Mining Techniques: A Source For Consumer Behavior Analysis**”, International Journal of Database Management Systems, September 2011.
- [4]. M. Alex Syaekhoni, Ganjar Alfian and Young S. Kwon, “**Customer Purchasing Behavior Analysis as Alternatives for Supporting In-Store Green Marketing Decision-Making**”, mdpi journal sustainability, 2017.
- [5]. Muqaddas Gull and Arshi Pervaiz, “**Customer Behavior Analysis Towards Online Shopping Using Data Mining**”, IEEE, 2018.
- [6]. M. N. Saroja, S.Kannan, K. R. Baskaran, “**Analysing the Purchase Behavior of a Customer for Improving the Sales of a Product**”, International Journal of Recent Technology and Engineering (IJRTE), Volume-7 Issue-4S, November 2018.
- [7]. Savi Gupta and Roopal Mamtara, “**A Survey on Association Rule Mining in Market Basket Analysis**”, International Journal of Information and Computation Technology, Volume 4, Number 4 (2014), pp. 409-414.

