



## International Journal of Control Theory and Applications

ISSN : 0974-5572

© International Science Press

Volume 10 • Number 30 • 2017

### Consumer Behavior Analysis using Big Data Analytic

K. Selva Kumar<sup>a</sup> Jasnoor Kaur<sup>b</sup> and S. Ramani<sup>c</sup>

**Abstract:** In any business association, varying consumer behavior is a huge challenge in sustainable growth of the business. In developing country like India, there is a shortage of resources to meet the basic requirements so there is a need to formulate and successfully implement strategies related to consumer behavior. Changing consumer behavior is an obstacle in the growth of business because it leads to heavy losses due to obsolete stock of the organization. Consumer behavior is very complex and sometimes unpredictable too. It is very often and not considered rational. A further challenge is that consumer personalities differ across borders and also between and within regions. Taste, behavior and preference of consumers cannot be ignored because consumers are the “Kings” of market. Consumer behavior is a complex, dynamic, multidimensional process, and all marketing decisions are based on assumptions about consumer behavior. Therefore, marketing strategies related to consumer behavior are made to beat cut throat competition in global context. In modern times, consumer behavior prediction is much essential for prosperity of the business. Its prediction and strategy formulation is a challenge for the management of any business organization. Only those organizations which formulate and implement consumer oriented marketing strategies, can survive in global competitive era.

**Keywords:** Business growth, Consumer Behavior, Global competition, Marketing strategy.

#### 1. INTRODUCTION

The study of the individual who wants to purchase certain product from shopping place, at the time of buying product, why he wants to buy it? This is very interesting concept to study. To study his mindset and converting the information into statistical format and analyzing their buying behavior.

The study can help firms and organizations to improve their marketing strategies by understanding the consumer's perspectives and issues such as:

1. The psychology of consumers that how he thinks, feel reasons and select between different alternatives.
2. The mindset of how the consumer is influenced by his or her environment.
3. The behavior of consumer while shopping or making other marketing decisions.
4. How customer motivation and decision strategy differ between products that differ in their level of importance or interest that they entail for the customer; and
5. How management can adjust and improve their marketing campaigns and marketing ideas to more effectively reach customer.

The question which arises in everyone's mind is what data is and why it is so important in analysis part? Data is basically refers to meaningless, raw symbols and figures. It has no meaning of its own. If in any ways, we can interpret the data then it can be considered an information for that data. Information has higher value than raw data as it has meaning associated to it. However, information may or may not be helpful. Knowledge contains higher value than information as it is a collection of information with a context. Knowledge can also be a thought of collection of information that is remembered. All the above described categories are a product of historical experience and processes and they help in answering certain queries of related context. The final but not the least category 'wisdom', it is the only category that deals with future. Wisdom is considered to have the highest value because it integrates vision and creativity.

Data refers to meaningless, raw symbols and figures. It has no meaning of itself. If data is interpreted in any way it can be considered as information. Information has higher value than raw data because it has meaning associated to it however information may or may not be helpful. Knowledge contains higher value than information because it is a collection of information with a context. Knowledge can also be thought of as a collection of information that is remembered (Bellinger et al., 2004). All the above described categories are a product of historical experience and processes and they help in answering certain queries of related context. The final category 'wisdom' is the only category that deals with future. Wisdom is considered to have the highest value because it incorporates vision and creativity. According to Bellinger et al. (2004) wisdom can only be created by entities that have souls and therefore computers are incapable of creating wisdom. This statement is particularly interesting in current context because Big Data is considered as the technology that in future will have potential to generate wisdom and generate insights and answers without having to ask questions.

### **1.1. Consumer Behavior**

The study of individuals, groups or organizations about their process of selecting, sequencing, using and disposing the products, services experiences or ideas to satisfy needs and impacts of these processes on the consumer and the society, refers as Consumer Behavior.

Behavior can concerned either with the individual or the group or with the firm. The product consumption is often becomes as much as important for the buyer, it may influence how a product is best positioned or how we can encourage increased consumption.

The behavioral analysis of consumer is to understand the process of their needs, if they can seize the behavioral characteristics and propensity to buy consumer will be able to retain existing customers and achieve customer relationship management purposes.

When consumer buys important products such as clothes, the decision-making process is more complex. In this type of decision-making, consumers seek information and evaluate different brands more thoroughly than in other types of decision-making situations. Consumers are the kings of markets. Without consumers no business organization can run. The activities of business rolls around with consumers and consumer satisfaction. Customer behavior study is centered on consumer buying behavior, with the customer playing the three dissimilar roles of user, payer and buyer. Research has revealed that consumer behavior is difficult to predict, even for experts in the field. The study of Consumer Behavior is quite complex, because of many variables involved and their tendency to interact with and influence each other. Consumer behavior is the study of when, why, how, and where people do or do not buy a product. It attempts to understand the buyer decision making process, both independently and in groups.

The study of analyzing the behavior of customer is based on its buying behavior, as customer playing three distinct roles of user, payer and buyer. Research has shown that analyzing the behavior of customer is difficult to predict, even for the suppliers. Relationship marketing is an influential asset for customer behavior analysis as it has a keen interest in the re-discovery of the true meaning of marketing through the re-affirmation of the importance of the customer or buyer. A greater importance is also placed on consumer retention, customer relationship with management, and personalization. Social functions can be categorized into social choice and welfare functions.

## 2. LITERATURE SURVEY

Several studies on consumer purchasing behavior had been presented and used in real Problem. Data mining techniques were expected to be more effective tool for analyzing consumer behavior. However the data mining methods had disadvantages as well as advantages. Therefore, it is important to select appropriate tool to mine database. The Junzo Watada and Kozo Yamashiro in their paper “A Data mining approach to consumer behavior” tried to improve data mining analysis by applying several methods including fuzzy clustering, principal component analysis and discriminate analysis. Many defects included in the conventional methods are improved in this paper. [2]

Prasanna Mohan Raj [3], studied the factors influencing customers brand preference of the economy segment SUV's and MUV's. Data collection was made through direct interaction and customer intercept survey using questionnaire. Descriptive analysis was used to transform data into understand format and factor analysis was used for identification of factors influencing customer preference. In light of study findings, the preference of a given brand can be explained in terms of six factors namely Product reliability, monetary factor, trendy appeal, frequency of non-price promotions offered, trustworthiness and customer feeling or association towards brand. There is need for marketers to take these factors into consideration when crafting product innovations in the SUV segment of Automobile market.

Social Media Analytics [4] enables organizations to utilize social media data for business insights and measure the effect of social media initiatives. Various benefits of social media analytics are - identify and address customer concerns to retain customer loyalty to products, understand new requirements and prospective customers, address competition and obtain customer feedback for proactive decision making. Various stages of social media analytics are: data collection, measurement - market sentiment analysis of products, data analysis and visualization - for better insights, innovative and Strategic business decisions.

B. Yıldız and B. Ergenç (Turkey) in “Comparison of Two Association Rule Mining Algorithms without Candidate Generation” Association rule mining techniques play an important role in data mining research where the aim is to find interesting correlations among sets of items in databases. Although the Apriori algorithm of association rule mining is the one that boosted data mining research, it has a bottleneck in its candidate generation phase that requires multiple passes over the source data. FP-Growth and Matrix Apriori are two algorithms that overcome that bottleneck by keeping the frequent item sets in compact data structures, eliminating the need of candidate generation. To their knowledge, there is no work to compare those two similar algorithms focusing on their performances in different phases of execution. In this study, they compare Matrix Apriori and FP-Growth algorithms. Two case studies analyzing the algorithms are carried out phase by phase using two synthetic datasets generated in order (i) to see their performance with datasets having different characteristics, (ii) to understand the causes of performance differences in different phases. Their findings are (i) performances of algorithms are related to the characteristics of the given dataset and threshold value, (ii) Matrix Apriori outperforms FP-Growth in total performance for threshold values below 10%, (iii) although building matrix data structure has higher cost, finding item sets is faster. [5]

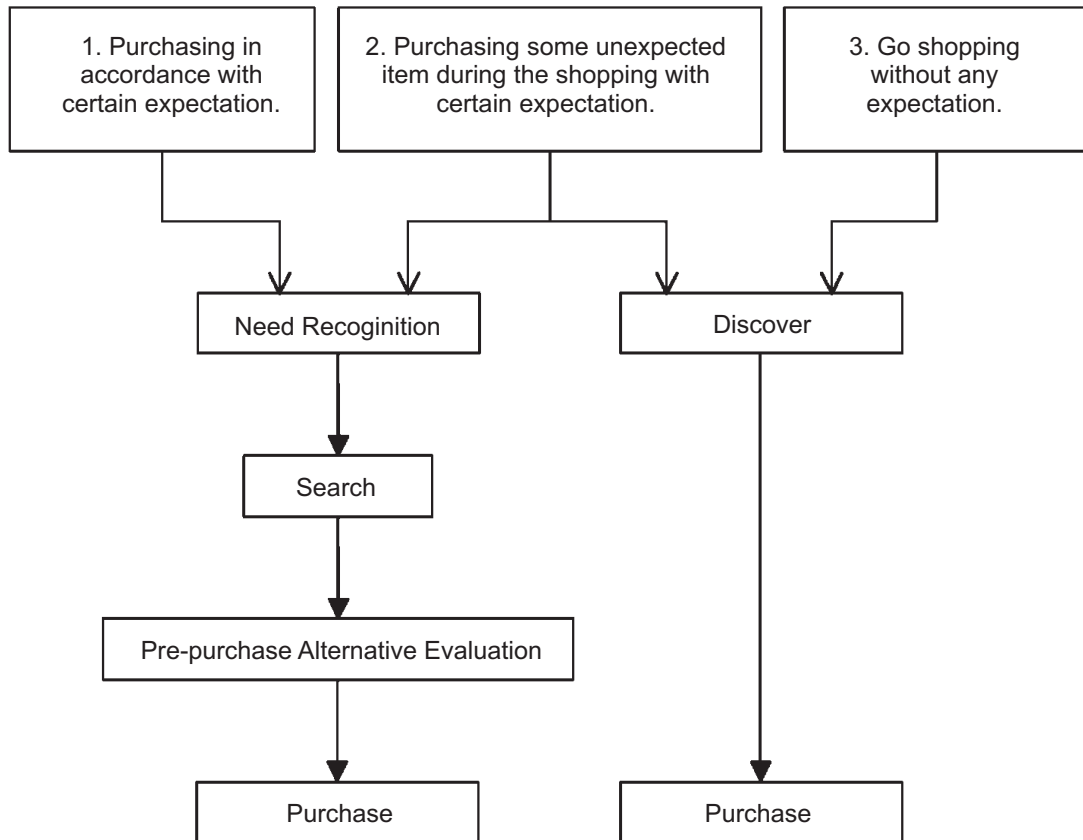
S. Vijaylaxmi, V. Mohan, S. Suresh Raju in their paper “Mining of users' access behavior for frequent sequential pattern from web logs” explains, in sequential pattern Mining comes in association rule mining. For a given transaction database T, an association rule is an expression of form X\_Y holds with confidence \_ % of transaction set T if \_ % of transaction set T support X UY. Association rule Mining can be divided in to two steps. Firstly frequent pattern with respect to support threshold minimum support are mined. Secondly association rules are generated with respect to confidence threshold minimum confidence. [6]The paper discussed the factors influencing pro environmental consumer behaviors and the policy implications of knowledge about these influences. They presented a conceptual framework that emphasized the roles of both personal and contextual factors and especially of their interactions. The practical usefulness of the framework is illustrated by evidence of the interactive effects of information and material incentives – typical interventions in the personal and

contextual domains, respectively [7]. A key difference between online and offline consumer behavior is that the online consumer is generally more powerful, demanding, and utilitarian in her shopping expeditions. As a result, customer loyalty on the Web is low overall (Morrisette et al. 1999); although, as in the physical world, it is influenced by the availability of good, relevant content in an enjoyable context (Rice 1997, Eighmey 1997, Eighmey and McCord 1998). The locus of power seems to be shifting from the vendor to the consumer (Raman 1997) who does not favor traditional advertising and promotions online (Maignan and Lukas 1997). Research has also shown that certain site features, such as the availability of a FAQ section or promotions at the Web store entrance, can influence both traffic on the Web site and overall sales (Lohse and Spiller 1998a and b) [8]. The paper proposed a system for distributed implementation using MapReduce Framework for C4.5 Algorithm along with customer data visualization. The author had introduced the implementation of C4.5 algorithm using MapReduce framework [9]. [10] In this paper, the author had tried to understand the trend of customers while using credit cards. So, the author had proposed a strategy where analyst can understand the customer's queries while taking credits cards or analyst would perfectly get to know how to deal with every type of customer.

### 3. METHODOLOGY

The flow tries to describe the decision-process behavior of customer. The process includes that when a customer is involved in purchasing goods, firstly he will go through the pre-purchase section, which includes need recognition, search for information and pre-purchase alternative evaluation.

The decision process of customers now-a-days varies with time and the factor which varies is shown below in the figure.



**Figure 1: Flow Diagram**

### 3.1. Data Analysis through Analytical Tools

Data analysis tools are used to sort through enterprise data in order to identify patterns and establish relationships. Analysis means to break a whole into its separate components for individual examination. Applying data analysis techniques can help virtually any business to gain greater insight into organizations, industry and customer trends. The below figure will explain the steps involved while performing Data Analysis.

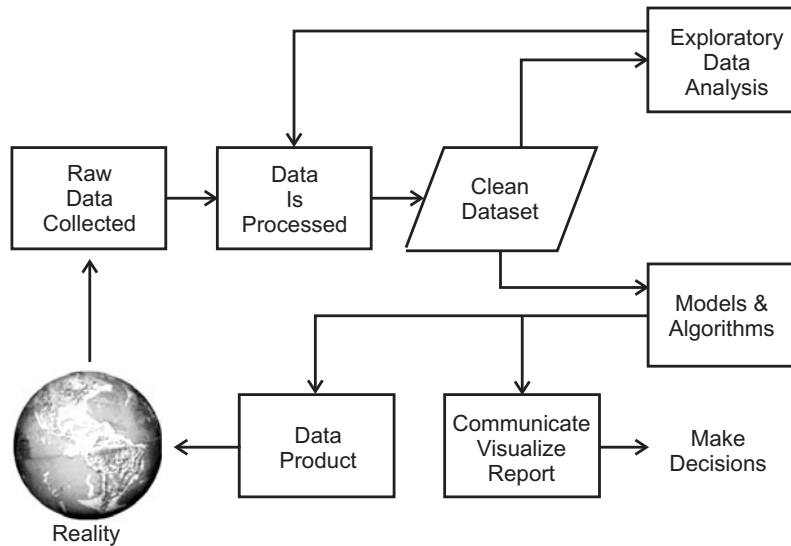


Figure 2: Data Analysis Process

Data Analysis is a process for attaining raw data and translating it into information valuable for decision-making by users. Data from various sources is gathered, reviewed and then analyzed to form some sort of finding or conclusion.

To analyze the data sets by using analytical tool such as Weka, Rapid Miner and Python packages such as panda, numpy and so on. By using these analytical tools, we can analyze the data and can come across to the conclusion about customer’s needs and try to fulfill their wishes.

1. Data Preprocessing
2. Data Cleaning
3. Data Migration using Analytical Tools

### 3.2. Data Preprocessing

In this module, the data sets creation is based on various factors. First step is to collect the data properly in the accordance to customer’s needs. The raw data are highly susceptible to noise, missing values and inconsistency, which can affect the quality of data. In order to improve the efficiency and quality of data, raw data is preprocessed as it transforms raw data into an understandable format.

### 3.3. Data Cleaning

The raw data, which is collected, needs to analyze according to its attributes. The data collected through questionnaire survey can be incomplete, noisy, inconsistent or relevant data may not be recorded due to misunderstanding. Data cleaning, routines work to “clean” the data by filling in missing values, smoothing noisy data, identifying or removing outliers, and resolving inconsistencies.

### 3.4. Data Analytic

Various tools have been used to analyze the data for better understanding of data, to get a clear picture about the behavior of the customer. Through various tools, we come to know what exactly the customer’s requirements and desire. Analytic lead a way or an organization to make a growth for their own business, to know more better what their users want from them. Analytics can leverage data and convert it to actionable information that can benefit organization. It basically enables business people to understand and make decisions on the basis of information which they have already and to make proper assumptions or predictions about their own users.

For our purpose, we have used analytic tools like Rapid Miner, python packages such as panda, numpy and so on. To get a more clear observation, we have collected data from various sources, then reviewed and then atlas, make analysis to form some sort of finding or conclusion. These tools can accelerate the building of complete analytical workflows- from data preparation to modeling to business deployment- in a single environment, dramatically improving efficiency and shorting the time to value the customer’s time and to get the things in lesser time. By proper understanding of data, then we can apply different algorithms to attain the predictions on the behavior of users from historical data. For our data sets, we have applied algorithms namely Naïve bayes Classifier, Decision tree and Regression Classifier.

### 4. STEPS OF ALGORITHM

1. Raw data is collected through questionnaires survey.
2. The collected raw data is processed and study thoroughly based on the target attributes and class variables.
3. Then the data which is preprocessed can be incomplete, noisy or inconsistent. Data cleaning, clean the data by filling in missing values, smoothing noisy data if found.
4. Exploratory data analysis is done to help the researchers to understand the need and desire of customers based on the historical data.
5. The analysis phase is divided into three parts:
  - a) Decision Making Process and Involvement
  - b) Individual Differences
  - c) Group Influences
6. The algorithms applies on the data to predict the user’s activities and makes assumption for the growth of organization.

$$p(a | y) = \frac{p(y | a)p(a)}{p(y)}$$

7. The results then make for visualize to communicate with users and to show their satisfaction with the products.

### 5. RESULT

The below table evaluates the mean and standard deviation based on the attribute of clothing items and varies according to the trust, security and privacy.

**Table 1**  
**Evaluation on Clothing Items**

<i>Attribute</i>	<i>Parameter</i>	<i>Trust</i>	<i>Security</i>	<i>Privacy</i>
Clothing Items	Mean	27.25	27	33.005
Clothing Items	Standard deviation	12.589	10.657	13.125

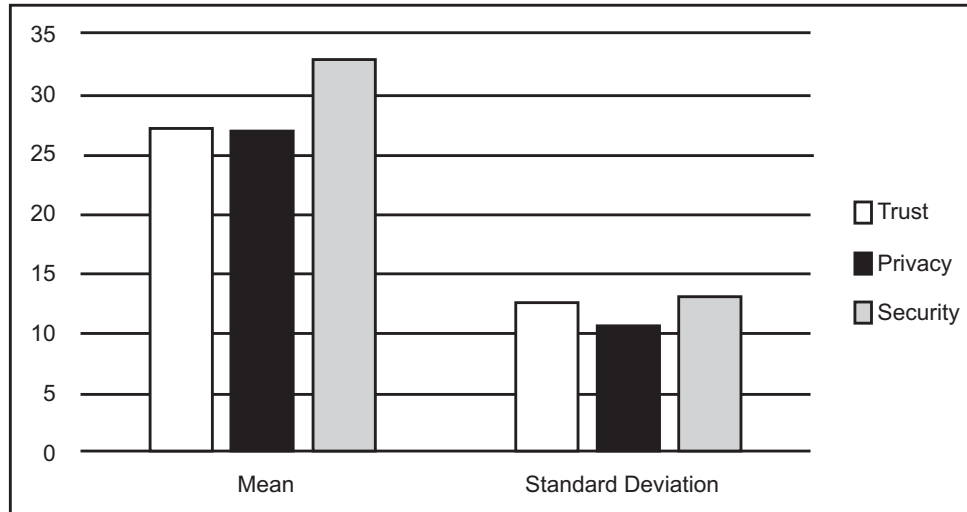


Figure 3: Bar Graph representing the factors

The graphs explain the information the user needs while they purchase any good and based on the data sets we get the below result.

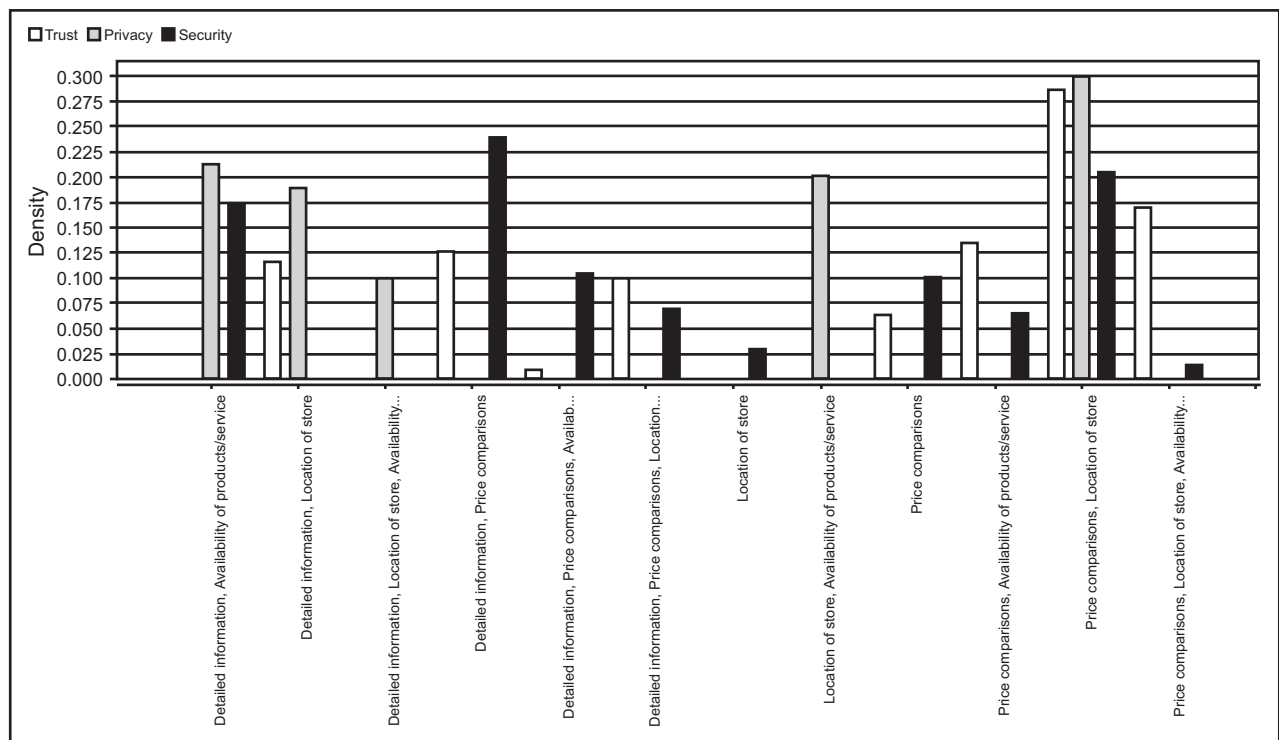


Figure 4: Information during purchase

The factor whether customer get influenced by others factors or not and which mode of method he prefer to choose for payment and on what factors he decide is given in the below graph, which explains the deciding factor which usually customer make, to help organization to grow their business.

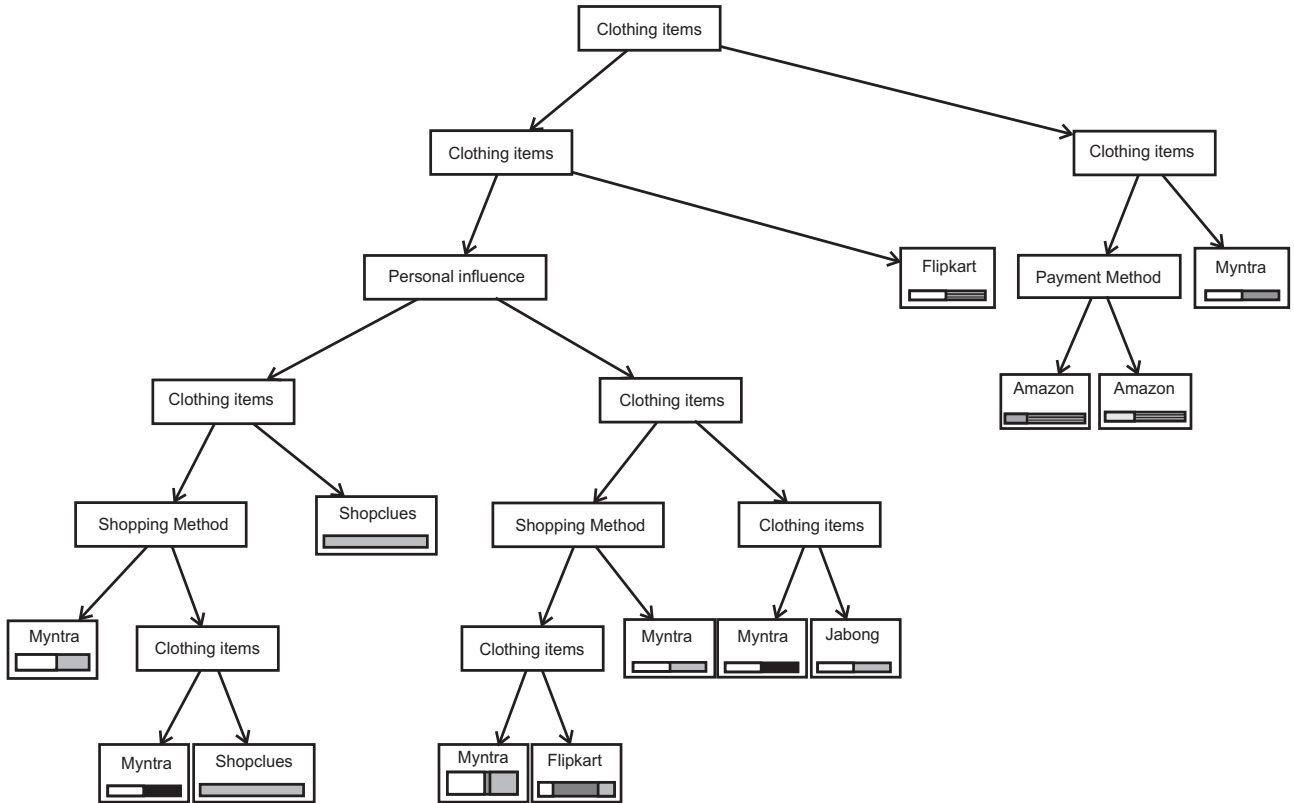


Figure 5: Decision Tree

Through analysis of data, the below graph predicts the type of clothes customer like to wear, it will help the shopper to keep themselves prepare according to the need of customer.

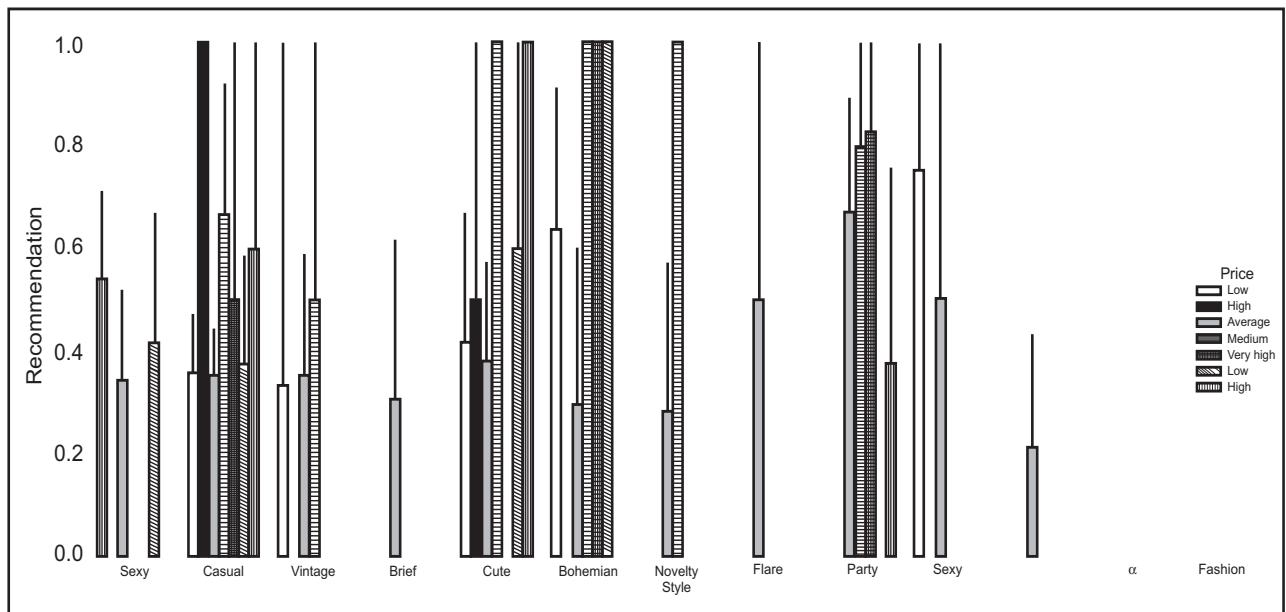


Figure 6: Preferable Clothes



## 6. CONCLUSION

Analysis of customer behavior enable companies to mend support of their customer oriented business processes, which aims to improve the whole performance of the enterprise. The paper centers on getting more customer satisfaction. The research paper involves in segmenting the customers and analyzing their behavior for further customer attraction, retention and development. As customer preferences evolve and options for where- and how - they shop expand, we all need a better understanding of who your customers are in order to stay competitive. The paper deals with the analysis of customer's behavior while shopping. Different analytic tools has been applied to fulfill the study of customer's behavior during shopping. The study included, different classification algorithms on the data sets such as naive bayes, decision tree and regression tree. The research tries to observe every possibility, to deal with every type of customers and tries to read their mindset and accordingly, satisfies their desires.

## REFERENCES

- [1] Raorane, A., & Kulkarni, R. V. (2011). Data mining techniques: A source for consumer behavior analysis. arXiv preprint arXiv: 1109.1202.
- [2] Watada, J., & Yamashiro, K. (2006, August). A data mining approach to consumer behavior. In First International Conference on Innovative Computing, Information and Control-Volume I (ICICIC'06) (Vol. 2, pp. 652-655). IEEE.
- [3] Raj, M. P. M., Sasikumar, J., & Sriram, S. (2013). A Study on Customers Brand Preference in Suvs and Muvs: Effect of Marketing Mix Variables. *Researchers World*, 4(1), 48.
- [4] Tan, W., Blake, M. B., Saleh, I., & Dustdar, S. (2013). Social-Network-Sourced Big Data Analytics. *IEEE Internet Computing*, 17(5), 62-69.
- [5] Yildiz, B., & Ergenç, B. (2010, February). Comparison of two association rule mining algorithms without candidate generation. In Proceedings 10th IASTED international conference on artificial intelligence and applications, AIA (pp. 450-457).
- [6] Vijayalakshmi, S., Mohan, V., & Raja, S. S. (2010). Mining of user's access behavior for frequent sequential pattern from web logs. *International Journal of Database Management System (IJDM)*, 2.
- [7] Sharma, S., & Lal, K. (2012). Changing consumer behavior—A challenge for sustainable business growth. *International Journal of Marketing, financial service and management research*.
- [8] Koufaris, M. (2002). Applying the technology acceptance model and flow theory to online consumer behavior. *Information systems research*, 13(2), 205-223.
- [9] Khade, A. A. (2016). Performing Customer Behavior Analysis using Big Data Analytics. *Procedia Computer Science*, 79, 986-992.
- [10] Roselin, R., & Hanupriya, C. (2014). Customer Behavior Analysis for Credit Card Proposers Based on Data Mining Techniques. *International Journal of Innovative Research in Advanced Engineering (IJIRAE)* ISSN, 2349-2163.
- [11] Satish, B., & Sunil, P. (2012). Study and Evaluation of user's behavior in e-commerce Using Data Mining. *Research Journal of Recent Sciences* ISSN, 2277, 2502.
- [12] Goel, A., & Mallick, B. (2015). Customer Purchasing Behavior using Sequential Pattern Mining Technique. *International Journal of Computer Applications*, 119(1).
- [13] Lin, C. F., Hung, Y. H., & Chang, R. I. Mining Customer Behavior Knowledge to Develop Analytical Expert System for Beverage Marketing. *INTERNATIONAL JOURNAL OF COMPUTER TRENDS & TECHNOLOGY*, 1(4), 579-584.
- [14] Yada, K., Motoda, H., Washio, T., & Miyawaki, A. (2006). Consumer behavior analysis by graph mining technique. *New Mathematics and Natural Computation*, 2(01), 59-68.
- [15] Turcinek, P. A. V. E. L., Stastny, J. I. R. I., & Motycka, A. R. N. O. S. T. (2012). Usage of cluster analysis in consumer behaviour research. In *Proceedings of the 12th WSEAS International Conference on Applied Informatics and Communications (AIC '12)* (pp. 172-177).

- [16] Calvert, G. A., & Brammer, M. J. (2012). Predicting consumer behavior: using novel mind-reading approaches. *IEEE pulse*, 3(3), 38-41.
- [17] Dhandayudam, P., & Krishnamurthi, I. (2013). Customer behavior analysis using rough set approach. *Journal of theoretical and applied electronic commerce research*, 8(2), 21-33.
- [18] Priya, S. L., & Varatharajan, S. (2013). Parallel algorithm based consumer behavior analysis for generating personalized ontology system. *International Journal of Management, IT and Engineering*, 3(5), 157.
- [19] De Nisco, and G. Warnaby, "Urban design and tenant variety influences on consumers' emotions and approach behavior," *Journal of Business Research*, In Press, 2012.
- [20] Dr.Raj Kumar., and Rajesh Verma., Classification Algorithms for Data Mining:A Survey, *International Journal of Innovations in Engineering and Technology (IJJET)*, August 2012..