❏   1818

Segmentation of Retail Mobile Market Using HMS Algorithm

Koyi Anusha, Yashaswini C, Manishankar S
Department of Computer Science, Amrita School of Arts and sciences, Mysuru Campus,
Amrita Vishwa Vidyapeetham, Amrita University, India

Article Info	ABSTRACT
	In the modern world of marketing, analyzing the trends in market is a key point towards to scope of improvement of any company. Considering the analysis of a retail market is highly challenging where market trends change very frequently based on customer needs and interest. Market segmentation is one of the approaches included in analysis of market trends which gives a diverse view of the market. The research here concentrates, especially on a case study based on fast moving consumable goods market and identifying market change patterns by applying a novel data mining approach. Data mining includes a wide variety of techniques and algorithm which can be effectively used in the process of market analysis. The research work carried out coins a new algorithm which combines various association rules and techniques, the HMS (Hybrid Market Segmentation) algorithm with some specialized criteria is used to support the market segmentation. The primary data needed for the analysis and operation are collected through a questionnaire based survey conducted on people from various demographic regions as well as various age groups. Used a quota based sampling approach for the research, The data mining approach here helps to study the large dataset collected and also to extract the useful information required to model the system. The system here is a learning system which improves the market segmentation functionality as data set improves, The paper implements a hybrid data mining approach which effectively segments the retail mobile market in to various customer and product groups and also provides a prediction and suggestion system for company as well as customer.

Corresponding Author:

Koyi Anusha,
Department of Computer Science,
Amrita Vishwa Vidyapeetham University, Mysuru Campus,
#114,7th Cross, Bogadi 2nd Stage, Mysuru-570026.
Email: chowdaryanoosha@gmail.com

1.   INTRODUCTION

In any retail market customer is king, understanding customer needs and interests is the most challenging task in the field of marketing. Market analysis mainly concentrates on areas like identifying the purchase pattern of a customer, extracting customer profile and suggesting products for customers, finding the product demand versus sales ratio. Retail marketing comprises the selling goods and services to consumers through distributed channels to earn more profit. Retail market varies a lot from wholesale in it is operational behavior, retail market concentrates on numerous products sold out into various customers in smaller quantities, whereas a wholesale market provides a bulk product stock sale to bigger organizations or industries. There are different forms of retailing such as shop based, e-commerce, direct marketing, etc. ecommerce concentrates on marketing and sales, which takes place through (B2C) business-to-consumer transactions in an online shopping portal or through mail orders. Retail customers are more demanding expected such as availability, value, choices, accessibility and pricing etc., in order to fulfill these requirements of the customer's retailer is forced to change their landscape very rapidly. To satisfy the

customer requirements the retailers are planning for the new strategies such as new business models, new communication, new operation etc.

Market segmentation is the process of categorizing the market into different customer groups and product groups, based on several preferences such as, education, income, region, country, etc. which are dependent and independent parameters affecting the changes in the market, market segmentation helps in understanding the customer needs and interests of an individual very efficiently for the decision making in the market, and also providing an exact picture on buying patterns of a customer. Market segmentation is classified into various type segmentations like demographic segmentation, Geographic segmentation, Behavioral segmentation, Psychographic segmentation, occasional segmentation, cultural segmentation [1].

Identifying the up-growing trends in the field of marketing is required in the business process, as well as Extracting knowledge, information from a huge customer dataset is very important in business for decision making and business process. Sales pattern in the inventory and forecasting the great potential such as decision making, market competition and strategic planning is an essential factor in the business process [2]. Demographic store which mainly concentrates on a particular segment of the globe and products that is particular to a demographic area. General store- a rural store that supplies needs in the local community. Supermarket is a self-service store which provides all varieties of a product under a single roof. Hypermarkets-provides a wide collection of products with low margins, the cost is comparatively less that other retail format [3]. E-Tailor customer shop and buys the product he or she wants through the internet. There are several factors which should be remembered by the retailer to increase their sales and few of the important factors as Unethical Consumer Behavior is the factor where in the customer buys the product and then returns it back after the product fulfilled its purpose. Customer service is the factor which helps to understand the needs of the customer and the retailer as to act according to it and fulfill all the requirements of the customer and analyze how to improve customer satisfaction [4]. Independent Retail Customer Services- to analyze how to extend a small retailer and to provide a good service. Male/Female shopping behavior-to analyze how differently, male and female shopping behavior varies How to cater the market, according to these categories and thus finding the dependent and independent variables in research on market segmentation. Data mining places a vital role in the analysis as well as segmentation process as it takes all these factors above mentioned as input to form an inference engine [5].

Data mining is the process of extracting data from large set of raw data and summarizing it into useful information for the further use [6]. The analysis step of KDD (Knowledge Discovery in Databases) process is data mining. It is an analysis of extracting large quantities of data from a previously unknown, interesting patterns such as Association Rule Mining (dependencies), Clustering (groups of data records) and anomaly detection (unusual records). Data mining involves six Common tasks: classification, clustering, regression, association rule learning, and anomaly detection. Classification is the task of grouping known structure to apply to new data. Clustering-it is the task of grouping similar structures to a single group without using known structures in data. Regression-it is the task of attempting to find a function in which data with less error. Association rule learning-it is also known as dependency modeling, it is the task of searching the relationship between the variables [7]. Anomaly detection-it is also known as outlier/deviation detection/change. It is the task of identifying unusual data sets, that may be interesting or data errors that required to further use.

Brito, Pedro Quelhas, et.al. "Customer segmentation in a large database of an online customized fashion business" this paper provides customer segmentation in fashion business based on customer preferences like age, gender etc. They had used two different techniques to group categories of same customers. One is clustering and second one is sub group discovery. Based on these methods organization can easily understand what type of products customer prefers and based on that organization can manufacture the products [8]. Aditya Joshiet et al. "use of data mining techniques to improve the effectiveness of sales and marketing" This paper proposes cluster association mining approach to classify associated patterns of sale and classify stock data. To implement the behavior they uses two phase, first phase divide the stock in three different clusters on the basis of sold quantities such as DS(dead stock), SM (slow-moving), FM(fast-moving).in the second phase they proposed MFP(most frequent pattern)algorithm which finds frequencies of pattern values of the corresponding items [9]. Migueis et al. "customer data mining for lifestyle segmentation" This paper provides life style segmentation based on life style of the customers. They had used VARCLUS algorithm to cluster the same group of persons. Still they had used hierarchical algorithm which combines divisible and agglomerative algorithms for clustering [10]. Noori, Behrooz et al. "An analysis of mobile banking user behavior using customer segmentation" taking a case study of an Iranian bank to find the customer usage of the online banking. Using a methodology of RFMD to find the recency, frequency of the customers and to identify purchased of different products through online mobile banking [11]. Baradwaj, B. K, et al. "Mining Educational Data to Analyze Students Performance" in the paper discussed current situation as well as future. They had used different data mining techniques to predict the

student performance in all activities such as exams and games etc. Mainly used ID3decision tree to find a best possible way to organize a learning set in questions. Data had collected from one university and stored information. Based on previous data system predicts the students' performance [12].

Liu, Xingli, and Huali Liu. "An Improved Apriori Algorithm for Association Rules" this paper explains about the improved apriori algorithm in association rule mining, this improved apriori algorithm as improved the efficiency and it establishes a new database to simulate applied experiment consisting of student achievement in computer programming course [13]. Zhou, Qishen et al. "Intelligent Data Mining and Decision System for Commercial Decision Making" this paper tells the importance of the information system and the database, which is very most important in the business for the decision making according to the rapid changes been considered in the business, it considers the most five themes to understand the competitive change in the business strategies such as leverage the franchise, intensify non-apparel, accelerate retail- led growth, invest in under- penetrated market and pursue operational excellence [14]. Hipp, Jochen, Ulrich Guntzer et.al. "Algorithms for association rule mining-a general survey and comparison" paper provides the comparison study between different algorithms used in market analysis. Used an algorithm, and analyzed their performance based on runtime experiments with maintaining threshold value. Paper concluded that in market analysis association rule works better than other algorithms [15].

2.   RESEARCH METHOD

Understanding the customer satisfaction ratio and buying pattern is the major concern of all the major retail marketing companies. Market segmentation and analysis plays a vital role in this context, the existing techniques in these fields did not give adequate care to analyze the interrelated parameters that plays a major role in queuing a solution to the market segmentation. There has been a paradigm shift that has taken place in the market, which gives the outlook only to the retail sector and alternating customer behavior. The major challenge that lies in this is coining an algorithm which mixes all aspects of a market factors as an input and generates an output for prediction or suggestion of market related deviations.

The initial step that has been taken towards the system framing was creating a user interface for both the customer and company to record their market related transactions. These inputs were preprocessed and stored into a database for the future usage of the inference engine. The inference engine functionality is bifurcated into two distinct but interconnected modules the suggestion and prediction system as well as the analysis engine.

Prediction and suggestion system works on the basis of the previously analyzed transaction details, which were provided by the data collector. This part has some integral algorithms proficient in handling the interrelated data from customer and company, generating a pattern for easier prediction and suggestion.

The architecture diagram denoted as Figure 1 aims to portray the interconnection that makes the underlying platform for the proposed system. It also shows the various components and the data flow that happens inside the system.
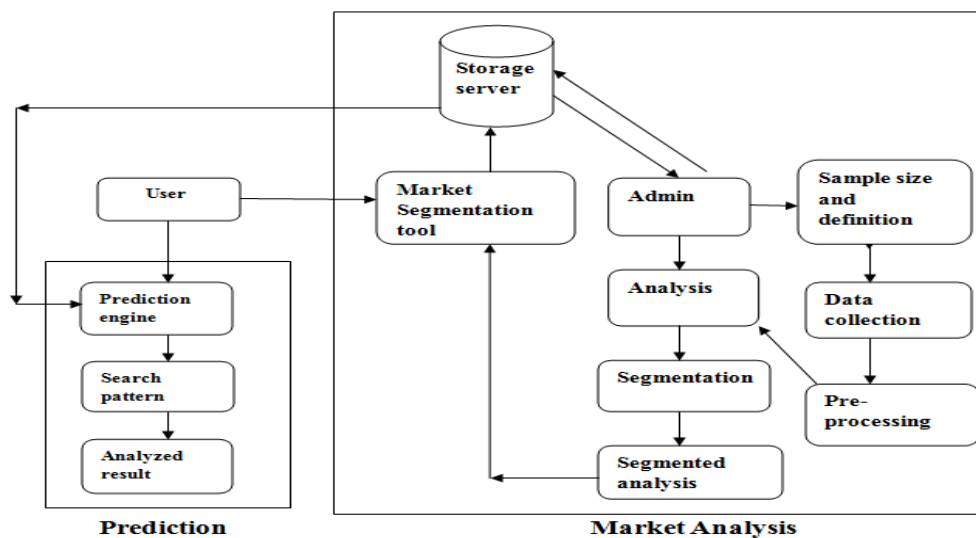

Figure 1. Architecture diagram

Analysis engine has a key role as it performs the segmentation part one of the prominent goals of the proposed system. Market segmentation tool is used to differentiate the various categories of customers that prevail in the retail market.

The administrator is responsible for the information regarding the previous transaction taken place with the same kind of products which the new customer is willing to buy, according to the parameters that have been considered in this work. We perform the pre-processing techniques on the top of the data in order to remove redundant, consistency and null data in the collected data to get exact data.

The next step in the process is to analyze the data with related information already gathered from the admin, perform different algorithms such as apriori algorithm and clustering techniques have been used to get common usage of products by the customer. Based on the analyzed information it is easy to get segmentation of the customer behavior through using different parameters. At the end segmented analysis will have the transaction information of previous customers, its directly stored information to the storage server. For new user Architecture helps to identify what type of products have to buy and it serves the requirements and satisfying the customer by serving them with best quality of services and products. The system provides a platform for the company to introduce their new products with the modified quality according to the customer preferences and likes.

The proposed system is been designed by using data mining technique such as Association Rule Mining which helps to identify the interest patterns between the variables in a huge database, the algorithm which is been used is the HMS algorithm which is helps mining the frequent item set, the algorithm helps to bring out the list of items been purchased by the old customer and helps to recommend the new customer to know what kind of products are been purchased, which helps the new customer to decide about the products buying. The other technique which is used is segmentation, Segmentation is the process where it helps the market to divide into a certain group or an individual, this division of customers helps the retailers improve their business and focus on the sales of the item in a profitable way. It also helps in understanding the individual or a group of similar liking customers and serves them according to their requirements and needs. There are many categorizing techniques in which we can segment the customers, in this work we have considered few of the parameters to segment the customer and fulfill their requirements and needs, the parameters which we considered are age, gender, occupation, income, education, behavior these segmentation help our work in segmenting what type of customers buy what kind of mobile phones and how to increase the sales of the product and also helps in modeling a new mobile phone.

2.1. Hybrid Market Segmentation Algorithm

HMS algorithm $(\sum D, \alpha, \beta, \pi, \Omega)$

$\sum D$ –Is the input data set

α-is the supplementing parameter 1

β-is the supporting parameter for α

π-is the similarity parameter for α and β for the data input $\sum D$

Ω -is the ratio of disparity value among α and β

The algorithm takes an input data set collected, and applies the four parameters α, β, π, and Ω in varying manner according to the rule for segmentation like demographic, Geographic, Behavioral, Psychographic, occasional, cultural segmentation and frame the parameter scale for each of the scenario.

The procedure is as follows by taking input and matching it with the test case segments and then providing a score value in each case. The final score value which is an aggregate of the individual scores helps in creating a comparison parameter which should be matched with the threshold obtained from the previously analyzed data.

Step 1 : input the data $\sum D$ from the storage sever

Step 2 : hierarchically order the segmentation rules with provided α, β, π, Ω and data input

Step 3 : define the test case with initial supplementing parameter α

Step 4 : classification based on the supplementing parameter is divided in to four classes C1, C2, C3, C4

Step 5 : C1- refers to the classification based on age group

Where α–is the age of customer

        β-is the various age group

        π-is the similarities in same age group ratio

        Ω-disparities in age group in purchase pattern

$|\pi| = \Delta \beta - \Delta \alpha$,where Δ is the deviation and $|\Omega| = \Delta \beta / \Delta \alpha$

$|R1| = |\pi| + |\Omega| + |\Delta \beta \text{ X } \Delta \alpha|$

$|R1|$ correspondence to the ratio of the segmentation

Step 6   : C2-refers to the classification based on the occupation

Where α-is the occupation of customer

β-is the various occupation group

π-is the similarities in the same occupation ratio

Ω-disparities in demographic group in purchase pattern of the customer

$|\pi| = \Delta\beta - \Delta\alpha$, where Δ is the deviation and $|\Omega| = \Delta\beta / \Delta\alpha$

$|R2| = |\pi| + |\Omega|_+ |\Delta\beta \times \Delta\alpha|$

$|R2|$ correspondence to the ratio of the segmentation

Step 7   : C3-refers to the classification based on the Qualification

Where α-is the qualification of customer

β-is the various qualification group

π-is the similarities in the same qualification ratio

Ω-disparities in behavioral group in purchase pattern of the customers

$|\pi| = \Delta\beta - \Delta\alpha$, where Δ is the deviation and $|\Omega| = \Delta\beta / \Delta\alpha$

$|R3| = |\pi| + |\Omega|_+ |\Delta\beta \times \Delta\alpha|$

$|R3|$ correspondence to the ratio of the segmentation

Step 8   : C4-refers to the classification based on the gender group

Where α-is the gender of customer

β-is the various gender group

π-is the similarities in the same gender ratio

Ω-disparities in demographic group in purchase pattern of the customers

$|\pi| = \Delta\beta - \Delta\alpha$, where Δ is the deviation and $|\Omega| = \Delta\beta / \Delta\alpha$

$|R4| = |\pi| + |\Omega|_+ |\Delta\beta \times \Delta\alpha|$

$|R4|$ correspondence to the ratio of the segmentation

Step 9   : Creating a classification ratio factor to perform the segmentation

$|R| = |R1| + |R2| + |R3| + |R4|$

Step 10 : Define a threshold ratio value which is the aggregate of all previously collected data ratio and each product group has a threshold value Δv.

Step 11 : Compare the $\sum R$ value with Δv value of each product class

Step 12 : Output the ratio value to the corresponding classes (C1, C2, C3, C4 etc.) This has the suitable match obtained in the comparison.

Step 13 : Generate the percentage based analysis graph plotting ratio values and threshold value.

3. RESULTS AND ANALYSIS

Table 1 gives the detailed information of customers with their occupation, qualification, age group, gender and income of the customer collected through multiple choice questions provided in the questionnaires. Dataset was collected through online web forms from different generations to identify their tastes and preferences about mobile retail market. Data set was rich with wide variety of data from different income groups as well as age groups. To segment the market mainly proposed algorithm considers the constraint like age group, gender, occupation, education and income of the family. Based on these constraints different purchasing patterns are generated to differentiate the retail mobile market.

Figure 2 provides the information regarding questionnaires asked in the survey to segment the customer buying patterns based on their choices Questions asked in the survey were mainly about mobile phones how customers select while purchasing mobile phones. Information got from various customers with different options and different preferences.

Table 1. Customer Details Group

S.No.	Attribute	Options	Values
1	Gender	Male	260
		Female	240
2	Age group	15-21	21
		22-30	167
		31-38	146
		39-45	92
		Above 45	67
3	Family Annual income	50000-100000	75
		100000-200000	200
		More than 200000	225
4	Education status	Computer science	149
		Commerce	174
		LLB	21
		MBBS	32
		Others	124
5	Occupation	Software engineer	60
		Bank manager	98
		Lawyer	23
		Doctor	32
		Student	82
		Others	205

Figure 2. Questionnaires to Collect Dataset

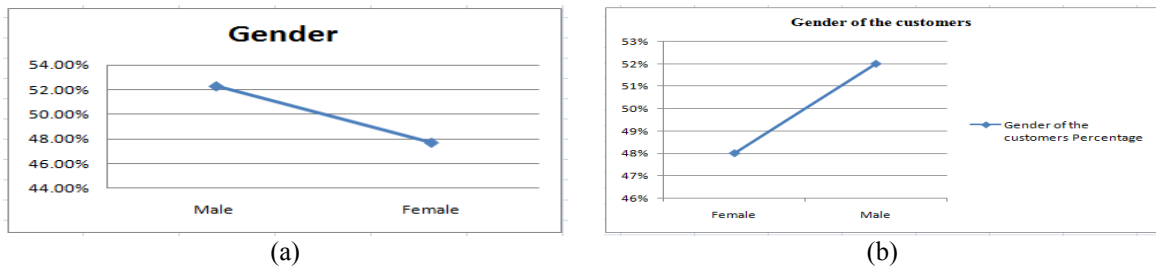| (a) | (b) |

Figure 3. Gender of the Customers

Figure 3 describes about gender category analysis which contains female and male customers in the retail market. Figure 3(a) is the status of the existing aprioiri algorithm based survey result it shows the result with the male of 53% whereas female as 47% and Figure 3(b) is the survey for the HMS alogrithm it gives the resut of male percentage is 52% where as female is 48% in the survey conduted for mobile phones.That means compare to females purchasing products males are purchasing mobile phones high for personal interactions. Both the algorithm results are similar and shows that male purchasing is the highest compared to the female.so both have similar true positives and false positives while taking gender as a constrain.

Figure 4 describes age group of the customers based on different scales. Mobile purchasing pattern will vary from one to another age group of the customers. Figure 4(a) is the survey of the existing apriori algorithm, it explains that the age group between 35 and 45 is the group where the purchasing of the mobile phones is at the highest and the age group between 18 and 25 purchasing is at the lowest. Figure 4(b) shows the result of HMS algorithm of 18-25 age group shows the percentage of 5% which is very low buying

patterns under this age group whereas 34-45 age group buys more mobiles as 36.7%. HMS algorithm shows the increased result by the 2% when compared to existing algorithm.
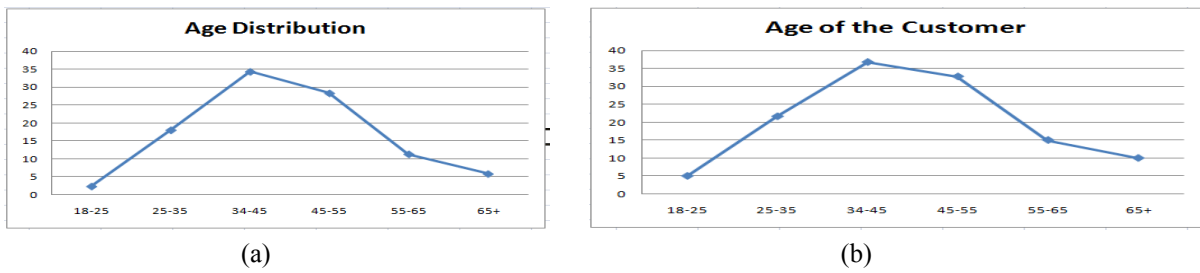

(a)   (b)

Figure 4. Age Group Of The Customers

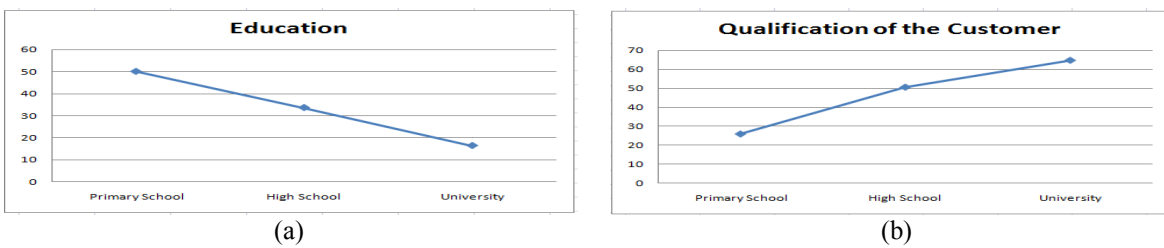(a)   (b)

Figure 5. Qualification of the Customers

Figure 5 describes the customer profile status based on their jobs, explains about the performance range of customer prefers what kind and how many mobile phones. Figure 5(a) is the result of the existing Apriori algorithm it gives the result between different levels such as primary school, high school and university and gives the result such that primary as the highest percentage and university as the lowest. Figure 5(b) gives the result of HMS algorithm in different attributes it shows the result of primary school attribute shows the least buying and the university attribute purchases the more mobile phones.

(a)   (b)

Figure 6. Occupation of the Customers

Figure 6 describes the ratio of the occupation of the customers, the graphs explains the details of the customer occupation category according to which the what kind of occupied customers buys what kind of mobile phones and what is the ratio of purchasing the mobile phones. Figure 6(a) is the result of the existing Apriori algorithm which includes different categories such as housewife, student, worker, engineer etc. according to the survey it shows that worker as the highest rate of purchasing the mobile phone and student is the least. Figure 6(b) shows the result of the HMS algorithm shows the result of the attribute tourism trade purchases the most and the housewife buys the mobile phones the least.

3.1. Extended features considered in HMS Algorithm

Income of the customer, number of mobile phones, brand of the mobile, purpose of the mobile phone and selection of the brand attributes are considered in the paper to obtain the better prediction and profile.
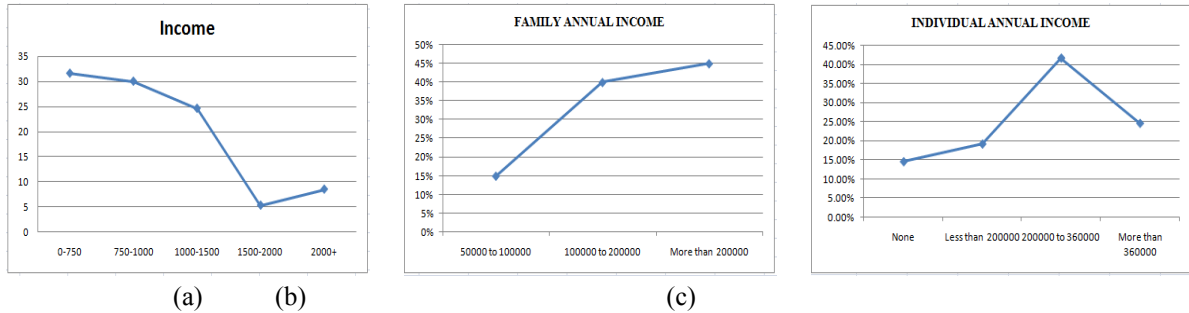

(a)   (b)   (c)

Figure 7. Incomes of the Customers

Figure 7 describes the income of the customer it shows the ratio of according to the income the customer purchases the mobile phones. Figure 7(a) gives the ratio of customer who's income varies and how their purchase the mobile phones. Figure 7(b) explains the ratio of the family income of the customer and the graph shows that the customer whose family annual income is more than 2, 00,000 purchases the mobile phones more and the customer whose family annual income is between 50,000 and 1, 00,000 purchases the least. Figure 7(c) explains the ratio of the individual annual income of the customer and the graphs shows that the customer whose individual annual income is between 2, 00,000 and 3, 60,000 purchases the mobile phones more and the customer whose individual annual income is none purchases the least. The previous research paper describes only the individual income but this paper gives the details of both the family annual income and individual annual income which gives the better understanding of the customer purchasing the mobile phones.
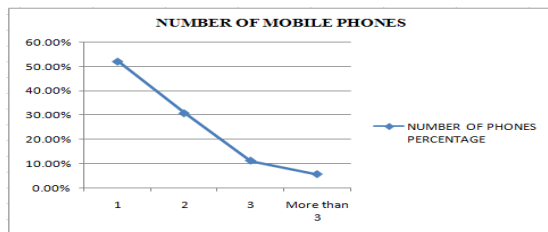

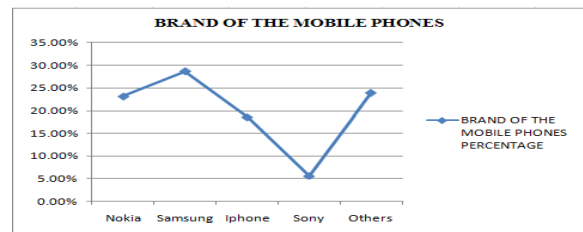Figure 8. Number of the Mobile Phones   Figure 9. Brand of the Mobile Phones

Figure 8 shows the information of the customers having the mobile phones. X-axis in the graph explains how many phones does the individual person consist of, whereas Y-axis explains the perentage of the mobile phones. Most of the customers are choosed only one mobile phone for usage with percentage of 52.20%. Less number of customers are selected more than 3 phones with percentage of 5.60%. Remaining customers choosed 2 and 3 mobile phones with percentage of 31% and 11.20% to use the mobile phone for their purposes.

Figure 9 tells about different categories mobile phones. X-axis in the graph tells about the different categories of the mobile phones such as Nokia, Samsung, IPhone, Sony and others, whereas Y-axis tells about the percentage of which category of mobile phone is been purchased. Considered the survey there is varying percentage of different categories of mobile phones being purchased in the case customers are mostly preferred samsung phones with percentage of 28.60% compare to other categories of mobile phones and sony is the least purchasing considered with the other categories of mobile phones.
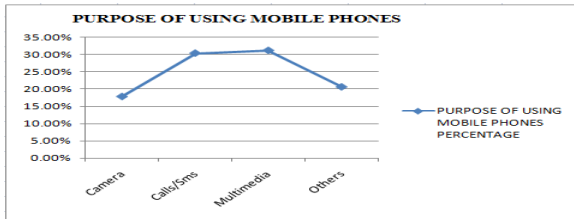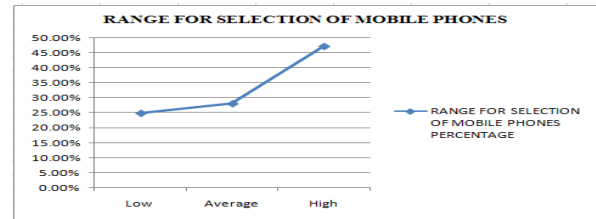
Figure 10. Purpose of using Mobile Phones

Figure 11. Selection of the Brand

Figure 10 explains for what purpose the mobile phones is been used, there are several purposes for which mobile phones is been used according to the individual customer behaviour, customer mostly uses mobile phones to different purpose like multimedia,camera,calls/sms and other uses. Considered the survey the purpose of using a mobile phone is multimedia with consisting of 31.20% compare to other options given in the survey to customers. Second selected option by customers was calls/sms with percentage of 30.40%, remaining options were consisting of less percentage compare to these options.

Figure 11 intreprets about the range for selection of mobile phones which includes low, average and high. Most of the customers preferred mobile range high cost with 47%. Remaining 27% customers preferred average cost mobiles and below 25% customers choosed low range mobiles to use their personal interactions. There is no much difference between low range and average range mobiles to purchase mobile phones.
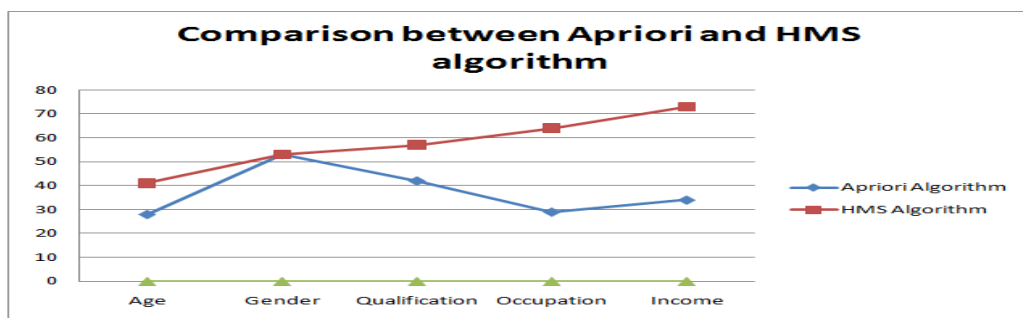

Figure 12. Comparisons between Apriori and HMS Algorithm

Figure 12 explains the ratio between Apriori algorithm and HMS algorithm. The above graph gives the detailed ratio between different categories such as Age, Gender, Qualification, Occupation and Income, the graph explains that the HMS algorithm gives the better result compared to the Apriori algorithm in the case of above categories taken into consideration.

4. CONCLUSION

Analyzing and predicting the purchase patterns and various constraints integral to retail mobile market using hybrid market segmentation (HMS) algorithm is being carried out the dataset was collected from various age groups. The results obtained supported the proposed methodology with an improved result compared to the existing algorithms used in segmentation. Improving the dataset population and adding more constraints can improve the performance of the algorithm. The algorithm can be generalized and used in various retail and wholesale market to get better segmentation and purchase patterns.

REFERENCES

[1] S. S. Gulluoglu, "Segmenting customers with data mining Techniques," *IEEE. Third International Conference*, pp. 154-159, 2015.
[2] B. M. Ramageri and B. L. Desai, "Role of data mining in retail Sector," *International Journal on Computer Science and Engineering (IJCSE)*, vol/issue: 5(01), 2013.
[3] A. A. Raorane, *et al.*, "Association rule– extracting knowledge using market basket analysis," *Research Journal of Recent Sciences*, vol/issue: 1(2), pp. 19-27, 2012.

[4]   M. J. Shaw, *et al.*, "Knowledge management and data mining for marketing," *Decision support systems*, vol/issue: 31(1), pp.127-137, 2001.

[5]   A. Parvaneh, *et al.*, "Combining data mining and group decision making in retailer segmentation based on LRFMP variables," *J Ind Eng Prod Res,* vol/issue: 25(3), pp. 197-206, 2014.

[6]   M. Y. Avcilar and E. Yakut, "Association Rules in Data Mining: An Application on a Clothing and Accessory Specialty Store," *Canadian Social Science*, vol/issue: 10(3), pp.75-83, 2014.

[7]   S. Suresh, *et al.,* "Application of Retail Analytics Using Association Rule Mining In Data Mining Techniques With Respect To Retail Supermarket," *IJEMR,* vol. 5, 2015.

[8]   P. Q. Brito, *et al.*, "Customer segmentation in a large database of an online customized fashion business," *Robotics and Computer-Integrated Manufacturing,* 2015.

[9]   A. Joshi, *et al.*, "Use of Data Mining Techniques to Improve the Effectiveness of Sales and Marketing," *International Journal of Computer Science and Mobile Computing (IJCSMC)*, vol. 4, pp. 81-87, 2015.

[10] V. L. Migueis, *et al.*, "Customer data mining for lifestyle Segmentation," *Expert Systems with Applications*, vol/issue: 39(10), pp. 9359-9366, 2012.

[11] B. Noori and W. T. Branch, "An Analysis of Mobile Banking User Behavior Using Customer Segmentation," *International Journal of Global Business*, vol/issue: 8(2), 2015.

[12] B. K. Baradwaj and S. Pal, "Mining educational data to analyze students' performance," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol/issue: 2(6), 2011.

[13] X. Liu and H. Liu, "An improved apriori algorithm for association rule," *TELKOMNIKA Indonesian Journal of Electrical Engineering*, vol.11, pp. 6521-6526, 2013.

[14] Q. Zhou, *et al.*, "Intelligent data mining and decision system for commercial decision making," *TELKOMNIKA Indonesian Journal of Electrical Engineering,* vol. 12, pp. 792-801, 2014.

[15] J. Hipp, *et al.*, "Algorithms for association rule mining—a general survey and comparison," *ACM sigkdd explorations newsletter*, vol/issue: 2(1), pp. 58-64, 2000.