



International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS)

www.iasir.net

Mining of Customer data in an Automobile Industry using Clustering Techniques

Archana Singh and Ajay Rana
Amity University,
Noida

Abstract— Customer Relationship Management (CRM) is a leading approach undertaken by the marketers in the process of retaining their customers. CRM approach is still new for the corporate managers, as in India the customers are still at the receiving end only. Data mining technology can be powerfully applied to the interdisciplinary research which brings out patterns and knowledge from the huge volume of data which in turn helps the Organizations in efficient decision making. Product and service based industries majorly deals with customers directly or indirectly. In this paper, it focuses on the study of type of customers clustered based on their patterns in an automobile sector. In this paper an attempt is made to segregate your customer with customer segmentation. This kind of knowledge will provide automobile business experts to develop and redefine their marketing strategies to update their CRM systems to attract more customers and increase the conversion rate.

Keywords- Data Mining; Consumer behavior; Clustering; Customer segmentation

I. Introduction

Market Pioneers realized by now that in the global and highly competitive market space, success rests on the firm's ability to attract, satisfy and retain its customers. To know your customer purchasing behavior has become the major requirements of the Organizations. Today's customer is very intelligent and aware. The customer gets knowledge from various sources like advertisements of the product, market campaign, references also from the review post on the social websites about the service and product by other customers. This knowledgeable and intelligent customer has increased the competition about the product/service in the market. Ngai(2008) in his research paper has done literature review on Data mining and customer relationships and mentioned about the framework of CRM. Hou and Tu(2008) explored customer relations and practices of business. Lifeng Tu[16] explains the degree of CRM and performance of new product. Cheng and Chen (2009) emphasized on customer segmentation value. To know the category of your Customer decides the further marketing strategy in attracting more customers [1]. This Customer entity in the datawarehouse is giving lots of data. In this paper, it focuses on customer buying behavior and type of customer which will help the Organization to redefine their market strategy to attract new customers and retain the old customers.

Data Mining is the technology which extracts knowledge from the data. Clustering (Han and Kamber, 2006) and Association rule mining are the techniques of data mining. Mining of data is a process, which consists of numerous steps such as integration of data from several databases or data warehouses, preprocessing of the data and induction of a model with a learning algorithm. The model is then used to identify and implement some decisions within the company. Transaction and survey data is collected to learn customer's behavior and buying patterns. Relevant and past information about the customer is collected and then analyzed. In Clustering, it categorizes the data of the same type. Data Mining has various clustering techniques. In this paper two Clustering algorithms are used, Two-step and K-Means.

The automotive industry in India is one of the sixth largest in the world [39]. The annual production as noted till 2011 is more than 3.9 million units. In 2010, India beat Thailand to become Asia's third largest exporter of passenger cars. In India, customer prefers that product which can provide easy, good service and the product should fit in their budget. There is lot of automobile companies manufacturing vehicles. Maruti Suzuki better known as Maruti is one of the established brands in India.

The paper is organized as follows : The Motivation behind and Research Methodology are discussed in section II. Approach and in Section III Empirical study and in Section IV Implications of the study finally, concluding the paper with future scope and Conclusion.

II. Related Work Done

An organization will have to change in order to truly support and foster CRM initiatives. The stages of transformation happened from Product to Customer. Organizations are shifting their emphasis according to the market trends and requirements. Transforming the business focus of an organization essentially means shifting the organization into the customer-focused paradigm. The business focus may shift from “How can we increase the sales of our products” to “What do our customers need, and how can we meet those needs?” Automobile industry is reaching out to people daily with its new four wheelers vehicles with innovative designs, giving more comfort, luxury and in affordable prices. Due to huge competition in the National and International market Manufacturers need profitable schemes for the customers. In CRM systems in order to support and enable meaningful customer dialogue at all points of contact. It is important to keep in mind that all the areas need to change in order to have an effective support CRM to enhance overall growth in an Organization. For the improvisations in existing CRM systems or building new systems data mining techniques are hugely applied.

Peng Gong[20] in his paper used case study of Shang Hua Carpet and applied improved apriori algorithm. Yaya Xie[33] explains the process driven architecture of building analytical CRM systems. Wanghualin [31] focuses on expatiating on the working methods of data mining technology . Zhang G.F[35] focused on Customer Segmentation and marketing strategy to be formed. .M.J Tarokh[18] aims on customer intelligence, collorative CRM(automobile industry). Jianing Qian[12] imparts customer knowledge in business growth. Also he has mentioned a case study of famous cosmetics retailing brand to show the role of data mining in Direct Marketing. Soumya Datta[27] shows the comparative study of statistical features of language. Jayanthi Ranjan[13] details about operational and analytical CRM which is giving knowledge of Organizational structure. Dezhen Feng[4] he discusses about customer value and the importance of CRM in the industry. This paper shows the effect of data mining in CRM. E.W.T. Ngai[17] shows a complete literature survey of CRM related research papers till the year 2006. Structure of CRM and various data mining techniques implications are well explained in the research paper. Jayanthi Ranjan[14] has done remarkable work in data mining and CRM . In this paper it refers to the business intelligence and factors affecting the business, organizational structure. Firms adopted BI. In research paper [3] data mining steps starting from data preprocessing ETL tools and building of data warehouse is mentioned.

RESEARCH METHODOLOGY

Data collection and analysis is viewed as a continuing and iterative process. Ideally, business decisions are refined over time, based on feedback from earlier analysis and decisions. Therefore, most successful analytical CRM research takes advantage of a data warehouse to provide suitable data. Considering the type of data required for the research and depending upon the purpose of the study, the data is collected from the trusted organization and by questionnaires from the customers. Many Research tools like Questionnaires, emails, and reviews post by the customer on social websites help to evaluate the CRM system.

III. APPROACH

An improved Customer service shall be the main driver for the Industry sectors to gain more profits that depends on the quality of their customer interactions to retain existing customer and win new ones. Data mining technology hugely applied in developing Intelligent CRM model [2][3][4]. The Data mining is the top level application research in database technology where it brings together artificial intelligence, machine learning and statistics [2][7].

Customer Segmentation

Grouping of Customers [11][35] according to their needs and response to product/service customers can be categorized and further services can be designed and planned accordingly. Clustering technique of Data Mining helps in categorizing customers according to the demand of product and need of features in that product. This kind of Customer categorization may term as Customer Segmentation. Each customer will be given the same list of products, but every customer has different needs and demands. Different opinions, interests will lead to different choices made by the customers. Hence, categorization of customers would become easy with varying parameters.

Customer Retention deals with the maintaining relationships with the existing customers. Repeated purchase of products and services by the customers makes them loyal to the Organization. Recommendation and Referencing the product and services to the other customers also add to the existing customer loyal and satisfied value. Invasion of media brings the popularity of good/bad of an Organization by the word of mouth, spreading the word about the product and services. Trust relationship between existing Customer and Organization once built can bring new customers as well more sales of products and in maintaining the same would result in fewer

costs than in investing cost in building trust relationship with new customers. Loyal or long term customer is most likely to buy the product from your Organization again in the near future.

The above can be achieved by Data Mining algorithms/techniques using the following approach.

- a) Using Data mining algorithms after knowing the interests of the existing customers under a group, individual customer buying behavior, interests, opinions about a particular product can be studied and a model can be build based on that.
- b) Using that data mining score model one can predict the future customer interest and buying behavior.
- c) Third step includes, analysis of each customer transaction data with the new score data mining model developed.

This analysis can be further use in determining and developing new strategies for more Profitability

IV. EMPIRICAL STUDY

The example is based on the customer data of famous Maruti Suzuki Brand, Maruti Suzuki India Limited (/maruti suzuki/), is popular as Maruti (Maruti Udyog Limited) an automobile manufacturer. As of November 2012, Maruti has 37% of the Indian passenger car market [37]. It has cars in various segments from the entry level 800, Alto, above average segment Ritz, Swift, Desire Luxury segment to the car in lower segment or Utility Cars like Eeco, Omni, and Multipurpose Vehicles like Ertiga and Sports utility vehicle Grand Vitara. They have various dealers in different parts of the Country (India). Maruti has various showrooms and workshops [38] located at Gurgaon (Haryana), Jaipur (Rajasthan) and Noida (UP). They are into sale/after sale service/exchange of Maruti cars.

The problem here, is focused on knowing your customer and your customer choices and preferences, few business objective is been targeted to propose a Customer Centric Business solution.

- a) Identifying prospective customer to become active customers.
- b) To identify Customer leaving the brand product
- c) To identify profitable (old and loyal) customers.
- d) To develop a marketing strategy to target the customer according to the preferences they have in the product.

For the analysis and research the data is not only collected from the registered customers of trusted dealer of Maruti but also from various sources like questionnaires, emails, social website review posts etc. The transactional data of customers their purchasing preferences, opinions about various cars in different models have been collected and analyzed.

For the segmentation of customer's two Clustering techniques of data mining Clementine is used. Two-step Clustering and K-Means [5]. With Two-step clustering is based on Agglomerative Clustering. It is an exploratory tool to generate automatic and hidden clusters. It handles large data files efficiently. After finding the number of clusters to be formed from this dataset collection of more than 600 customers another popular technique of data mining K-means is used. K-means groups the clusters on the basis of Euclidean distance. Car manufacturers need to be able to appraise the current market to identify the likely competition for their vehicles. If cars can be grouped according to available data, this task can be largely automatic using cluster analysis.

Data Analysis

With 12 input values including Demographic as well Customer preferences data has been collected. The attributes include Age, Variant Name, Expectations, Service Comfort, Performance, Mileage, Space Economy, Value for Money, Exchange_of_Car , Occupation, Source of Enquiry. The data has been collected from various sources like Emails, Questionnaire, direct from the dealer about his customer's transaction detail. Around 631 record sets have been used for analysis. There are various types of variant Car names which are categorized into 9 types. The age of Customer ranges from 22to 65. Using the data mining Clementine tool, Crosstabs Correlation description between Age and Variant names has been found.

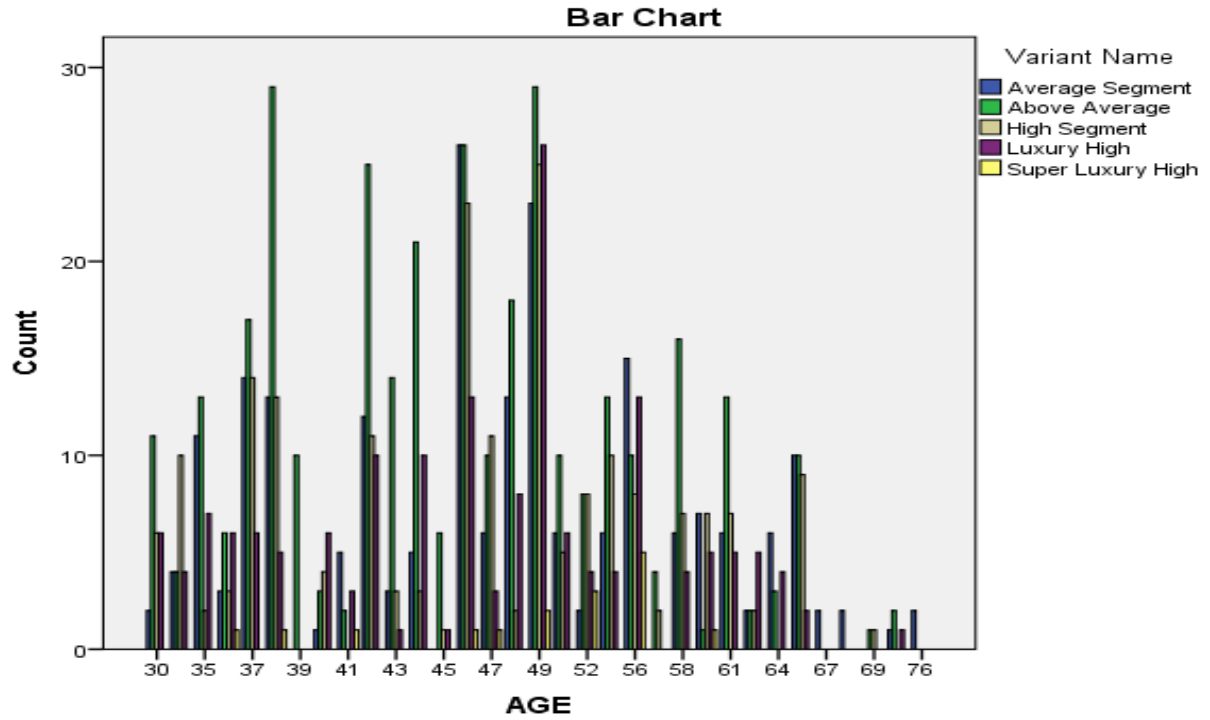


Figure 3 :Crosstabs Correlation between Age and Variant segments

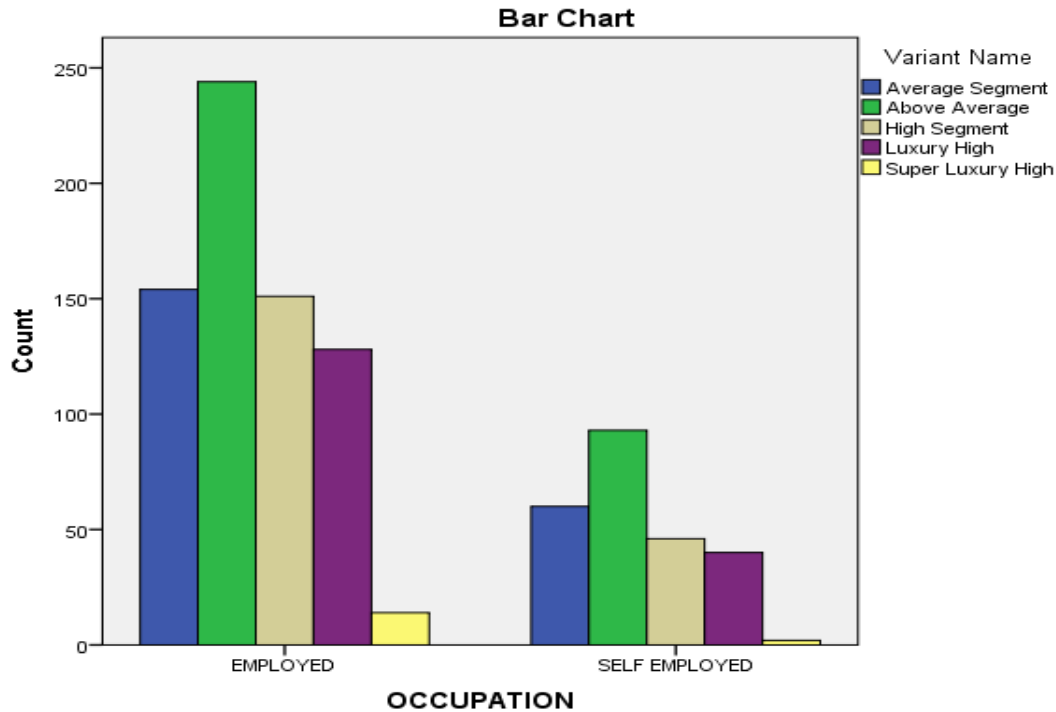


Figure 4 : Crosstabs Correlation between Occupation and Variant Segments

The Crosstabs technique is used and it has been found that Figure 3 shows crosstabs Correlation between age of Customers and different Variant/Vehicle Segments. Figure 4 shows Correlation between Occupations and different variant segments. Variants or Cars are segmented into 5 groups Super Luxury Segment, Luxury Segment, High Segment, above Average and Average Segment Cars. / Variants.

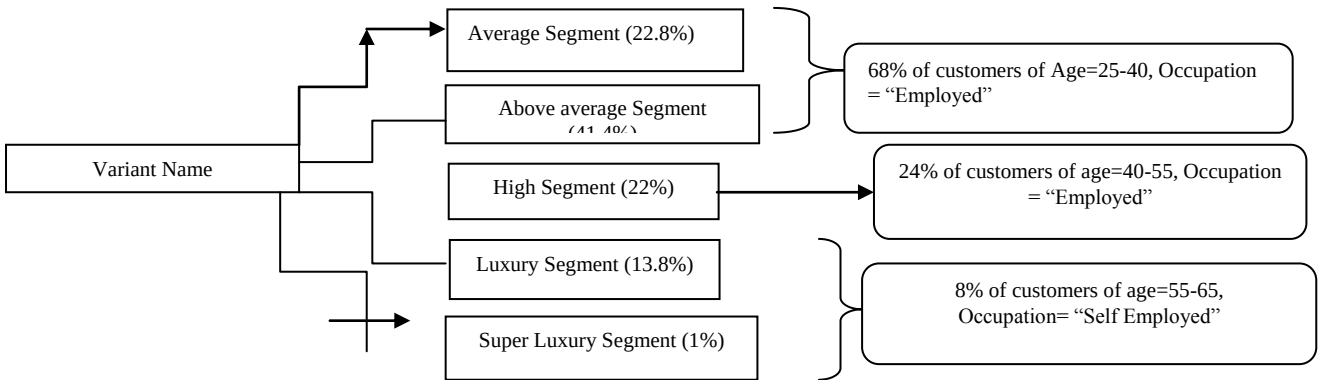


Figure 5: Correlation helps the Organization to know the demand of Cars according to the age and Occupation

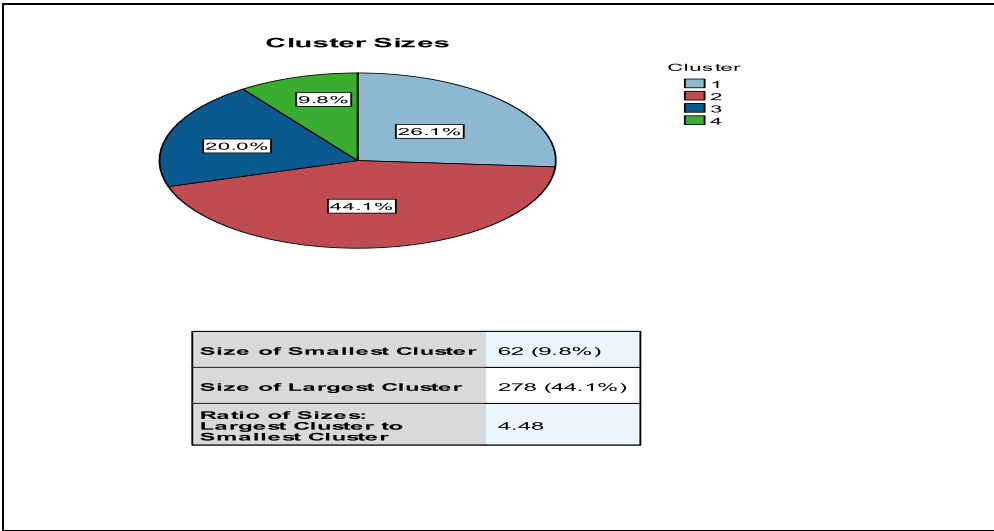


Figure 6(a) : Cluster Sizes and ratio between Largest and Smallest Cluster

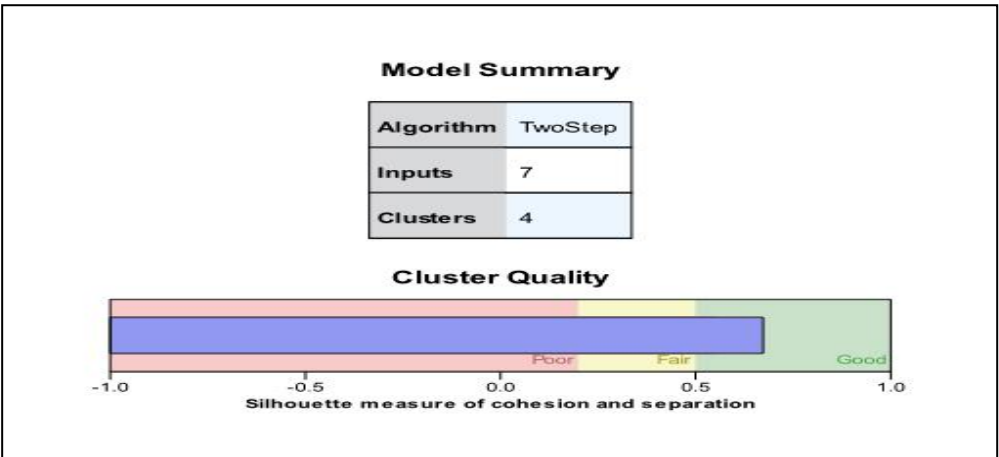


Figure 6(b): Model Summary of Cluster Analysis

With the technique of two step cluster it is predicted that 4 Clusters are formed. Figure 6(a) shows the Cluster sizes and ratio between the largest and smallest cluster and in figure 6(b) shows a summary model with 7 inputs, 4 clusters are formed .The bar shows the quality of a cluster which is based on Silhouette measure of cluster cohesion and separation the shaded progressive bar indicates that poor, fair and good (based on Kaufman and Rousseeuw, 1990) interpreting cluster structures. Here, it indicates a fairly good cluster structure means good quality of clusters.

Table 1:iteration History^a

Iteration	Change in Cluster Centers			
	1	2	3	4
1	9.239	9.356	6.122	2.735
2	2.300	1.379	.705	.551
3	.344	.534	.000	.000
4	.000	.000	.000	.000

Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is .000. The current iteration is 4. The minimum distance between initial centers is 16.941.

Table 2: Number of Cases in each Cluster

Cluster	1	120.000
	2	44.000
	3	175.000
	4	292.000
Valid		631.000
Missing		.000

No. of records in each cluster

Table 3:ANOVA

	Cluster		Error		F	Sig.
	Mean	df	Mean	df		
	Square		Square			
AGE	16107.297	3	7.830	628	2057.037	.000
Variant Name	2.192	3	1.835	628	1.195	.311
EXPECTATIONS	1.039	3	.757	628	1.372	.250
Exchange	.201	3	.247	628	.814	.486
Enquiry Status	.659	3	.341	628	1.934	.123
VALUEFORMONEY	1.950	3	.742	628	2.628	.049
Source of Enquiry	2.002	3	2.825	628	.709	.547
OCCUPATION	.246	3	.194	628	1.269	.284

The F tests should be used only for descriptive purposes because the Clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal

After identifying and filtration of parameters which are positive with the other parameter and shows the data highly Coorelated with that parameter. Then, the required parameters are filtered and notice that each parameter is showing different aspect of customer buying behavior. Now, to group those customers according to their behavior and other details.

Interpretation of RESULTS

From the Table 1 , it is interpreted that Convergence is achieved during the iteration process. Table 2 shows that number of records or cases in each cluster. Table 3 is indicating the Anova table shows that out of 12 variables which variables are significantly different from each other across 4 clusters. The Significance column indicates that customer preferences in Maruti Products benefits are Mileage, Value for Money, Performance, Space Economy, Comfort and Age are significant at level at 0.05 level or can say their value $p < .05$. The other variables are not significant due to their greater probability values.

By using above two Clustering techniques, Four Clusters are formed.

Cluster 1 : These Customers Expect “SERVICE COMFORT ” from the Vehicles of Maruti. For them Comfort, Mileage, Performance, Space Economy are Important and they don’t want to change Maruti Cars and their source of Enquiry is “Compaigh/Events” and their Enquiry Status is “Active” that means Maruti cars fulfill their expectations. These customers are the beneficial to the BRAND Maruti as Segment vehicles has better Space. They are Satisfied Customers in category High Segment Cars

Cluster 2 : These Customers Expect “ABOVE AVERAGE” from the Vehicles of Maruti. they can change Maruti Cars for any other car with better looks and their source of Enquiry is “INBOUND CALLS” and their Enquiry Status is PASSIVE that means. They are not satisfied Customers. The Company should form a policy for the conversion from Passive to active customers in category Very High Segment Cars.

Cluster 3: In this cluster, Customers Expect “VALUE FOR MONEY” from the Vehicles of Maruti. For them Comfort, Mileage, Performance, Space Economy are very Satisfactory and they don’t want to change Maruti Cars and their source of Enquiry is “SHOWROOM VISITS ” and their Enquiry Status is “INVOICED” that means Maruti cars fulfill their expectations . These customers are Active and Highly Satisfied Customers in category Average segment Cars

Cluster4:

The customers under this group don’t want to change Maruti Cars and their source of Enquiry is “Anytime Maruti” and their Enquiry Status is “invoiced” that means If Maruti cars fulfill their expectations. These Customers would go for MARUTI BRAND only. In Cluster4 these customers are more **Old and loyal** to the BRAND MARUTI in category Above Average Segment Cars

Customer Segmentation Analysis

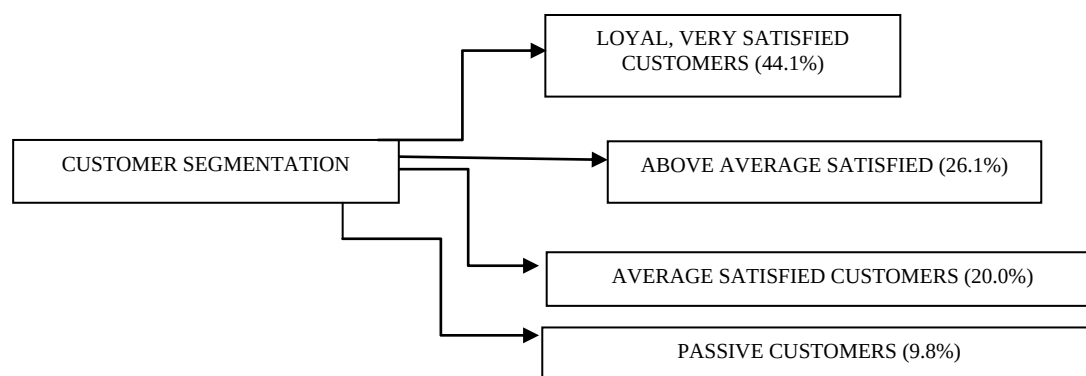


Figure 7: Segmentation of Customers

V. IMPLICATIONS OF THE STUDY

In this research paper, the study shows that Data Mining techniques in building customer centric solutions in Indian Organization like automobile sector gaining popularity and data results, analysis can be used for knowing your customer and its preferences.

With vast variety of products available in the market, the customers are left with many choices to make among the brands and within different variety of products of different segments. The proposed study identifies the type of customer and will be able to examine the relationship between customer relationship management and data mining techniques. Also, with the impact data mining techniques in CRM shows that pattern mining, feature extraction from the will help the management to build strategy to retain , attract new prospective active customers .

The Goal of Customer centric analytical model was to identify the Active , loyal customers those who directly purchases the variant/product as well as contributes in the Organization by referring to other customers, by posting their positive responses / review about the product and service on the various social websites. From the proposed model cumulative customer cost (product, service, recommendation, other hidden costs) is calculated and customer retention value can be computed with that rate can also be used for further decision making process regarding customers.

For Customer segmentation the adoption of parameters (age, Variant name, Expectations, Exchange, Enquiry status, value for money, source of enquiry, occupation) categorizes customers into loyal very satisfied, above average, average and passive customers. This segmentation can lead to better conversion rate from passive customers to active customers. Here , source of enquiry tells me if is anytime maruti and enquiry status is active then he is the most loyal customer active is judged as a measure of number of post, reviews on the social websites about the product and service.

Future Scope of the Paper

This paper can be further extended to know the significant and influential features affecting the sales of the Organization, keeping customer choices in mind.

VI. CONCLUSION

It is been observed and a well known fact that research and data analysis done using data mining techniques helps the Organization for better decision making and strategy planning. For the growth of any product and service oriented industry it is important to keep bringing new policies to attract new and to retain existing customers is the big and never ending challenge. In this research paper, a new Customer Centric Analytical model has been proposed using data mining, computing active customer retention value and the parameters affecting this value. An empirical study upon Automobile industry has been discussed, data analysis, result interpretation thus providing business solution to the Organization to improve prospect customer.

REFERENCES

- [1]. Berson, A., K. Thearling, and J. Stephen(1999), "Building Data Mining Applications for CRM", USA, McGraw-Hill .
- [2]. Dezhen Feng1 Zaimei Zhang1 Fang Zhou2 Jianheng Ji2(2008) , "Application Study of Data Ming on Customer Relationship Management in E-commerce", published in IEEE digital explorer
- [3]. Francesco Bonchi, Carlos Castillo, Aristides Gionis, and Alejandro Jaimes(2011). Social Network Analysis and mining for Business Applications. ACM Transactions on Intelligent Systems and Technology.
- [4]. G. Holmes; A. Donkin and I.H. Witten (1994). "Weka: A machine learning workbench".

- [5]. Giudici, P., Applied Data-Mining(2003): Statistical Methods for Business and Industry. West Sussex, England: John Wiley and Sons .
- [6]. [7] Hai Ying Xie, Wen Jing Tang(2009). Application Research of Data Mining in Travel Agency's Customer Relationship Management Second International Workshop on Computer Science and Engineering.
- [7]. Han, J., and M. Kamber, "Data Mining: Concepts and Techniques" New York: Morgan Kaufman (2000).
- [8]. Hughes A.M, "The Complete Database Marketer: Second-generation Strategies and Techniques for Tapping the Power of Your Customer Database", McGraw-Hill, 1996
- [9]. [9] Jhilmil Jain, Ismail Ari, Jun Li(2008). Understanding the Challenges Faced During the Management of Data Mining Models. 2nd ACM Symposium on Computer Human Interfaces for Management of IT (CHIMIT) .
- [10]. IndiaPRwire(2012) "India's Rapidly Evolving Enterprise Mobility Market To Touch USD 1 Billion By Mid-2015: Zinnov Study".
- [11]. Janny.C.Hoekstra, Eelko.K.R.E.Huizingh(1999) "The Lifetime Value Concept in Customer-Based Marketing", Journal of Market Focused Management [M], 3, 257-274.
- [12]. Jianing Qian et al (2011) "Application of data mining techniques in customer relationship management" published in IEEE digital explorer.
- [13]. Jayanthi Ranjan et al (2008) "Distinguishing operational customer relationship management and analytic customer relationship management-an organizational perspective" published in International Journal of Logistics Economics and Globalisation, vol1, Nos3/4.
- [14]. Jayanthi Ranjan (2008) "Implementing Business Intelligence in Indian firms: prospects and challenges" published in International Journal Indian culture and Business Management, Vol1, No. 3.
- [15]. Jayanthi Ranjan(2010) Integrating analytical Customer Relationship Management with Customer Relationship Management in organizations. International Journal of Innovation and Learning. IJIL, Inder Science Publishers. Vol 8, No 4.
- [16]. Jayanthi Ranjan Vishal Bhatnagar Raghuvir Singh(2011) Analytical customer relationship management in insurance industry using data mining: a case study of Indian insurance company. International Journal of Networking and Virtual Organizations (JUNVO) Volume 9 - Issue 4 -2011 pp 331-366. Inderscience publishers. http://www.inderscience.com/search/index.php?action=record&rec_id=43803&prevQuery=&ps=10&m=or
- [17]. Lifeng Tu(2011) "the effects of customer relationship management on new product design" published in Fourth International Conference on Business Intelligence and Financial Engineering(IEEE).
- [18]. Li Xiu b, Ngai, D.C.K. Chau (2008) "A literature review and classification published in expert systems and applications" published in Expert systems and its Applications, Elsevier, volume 36, 2008
- [19]. M.J Tarokh et al (2007) "Intelligence CRM : A Contact Center Model" published in IEEE.
- [20]. [17] Mingjun Song, & Sanguthevar Rajasekaran(2007) A transaction mapping algorithm for frequent item sets mining. IEEE Transactions on Knowledge And Data Engineering,
- [21]. Ranjan J et al(2011) "Multidimensional Data Modelling and Business Intelligence An Approach for Better Analysis" published in the proceedings of the International Conference on Facets of Business Excellence- Leveraging Information Technology for Strategic Advantage" pp 67-76.
- [22]. Peng Gong et al (2007) "The Application of Improved association rules Data Mining algorithm Apriori in CRM" published in IEEE digital explorer.
- [23]. Singh A, Ranjan Jayanthi et al(2011) "Application of Data Mining Techniques in Social Media Analytics" published in the proceedings of the International Conference on Facets of Business Excellence- Leveraging Information Technology for Strategic Advantage" pp 1-12.
- [24]. Singh A et al (2011) "Online mining of data to generate Association Rule Mining in Large Databases" published in IEEE.
- [25]. Singh A, Jayanthi Ranjan et al(2012) "Enterprise Mobility in Higher Education Universities: Data Mining Perspective" published in the proceedings of FOBE'2012.
- [26]. Smith, Sylvana (2003) SAS Institute white paper. SAS @ Banking Intelligence Solutions for Customer Relationship Management.
- [27]. Shuxia Ren, Qiming Sun and Yuguang Shi. Customer Segmentation of Bank based on Data Warehouse & Data Mining. Information Management and Engineering (ICIME), 2010 The 2nd IEEE International Conference
- [28]. Soumya Datta , Dr Sudeshna Sarkar (2008), "A Comparative Study of Statistical Features of Language in Blogs-vs-Splogs[]" " ACM, Singapore
- [29]. Su Xinning et al(2006), "Data Warehouse and Data Mining", Tsinghua University Press, pp. 43-96
- [30]. Vijay T. V, Singh A and Dubey G(2012) "Mining Queries for Constructing Materialized Views in a Data Warehouse" published in "Advances in Computer Science Engineering and Applications, AISC 167, pp 149-159.
- [31]. Wang Ke Nan et al(2008) "Apply of Data Mining Technology in CRM" published in the Changchun Taxation College Research Project and IEEE.
- [32]. Wanghualin(2010) "Data Mining and its applications in CRM " published in Second International Conference on Computer research and Development(IEEE).
- [33]. W.C Hair, J.F, et al(1998) "Multivariate Data Analysis". 5th edition. USA, PHI, Inc.
- [34]. Yongqiang Chen et al(2005) "Study on Data Mining Application in CRM System Based on Insurance Trade", ICEC'05, August 15-17, 2, Xi'an, China.
- [35]. Zhang G.F. , Jin G.D(2006) "Research on Customer Segmentation theories and their application Strategies", Journal of College of Mechanical Science and Engineering, pp.101-104.
- [36]. Zhou Jianhong et al(2008) "Design of Customer Relationship Management Systems Based on Data Mining" published in IEEE and International Symposium on Electronic Commerce and Security
- [37]. http://en.wikipedia.org/wiki/Maruti_Suzuki
- [38]. <http://www.vipulmotors.com/home.aspx>
- [39]. http://en.wikipedia.org/wiki/Automotive_industry_in_India