Proposed analytical customer centric model for an automobile industry

Archana Singh* and Ajay Rana

Amity University,
Sector-125, Noida(U.P), India
Email: archana.elina@gmail.com
Email: ajay_rana@amity.edu
*Corresponding author

Jayanthi Ranjan

IMT Ghaziabad,
Raj Nagar Hapur road, Sector 10,
Raj Nagar, Ghaziabad(U.P), India
Email: jranjan@imt.edu

Abstract: Research practitioners have contributed extensively in maximising the customer value using different models. It has been realised that customer lifetime models are industry dependent. The analysis of customer lifetime value is done with different customer value models. Recency, frequency, and monetary (RFM) model is traditional used for market segmentation to detect patterns customers/buyers. In this research paper, a new customer centric analytical model has been proposed for an automobile industry. The paper explores new dimensions in the proposed model to determine the customer value and further by using data mining approach helps the management to build new strategic customer relationship model. The data mining technique association rule mining, direct hashing and pruning is used to generate rules and market patterns for the market management which in turn facilitates in effective and efficient decision making.

Keywords: customer lifetime value; CLV; data mining; segmentation; recency, frequency, and monetary; RFM; association rules.

Reference to this paper should be made as follows: Singh, A., Rana, A. and Ranjan, J. (2015) 'Proposed analytical customer centric model for an automobile industry', *Int. J. Data Mining, Modelling and Management*, Vol. 7, No. 4, pp.314–330.

Biographical notes: Archana Singh completed her MTech (Computer Science and Engineering) from I.E.T.E New Delhi, a Masters' in Computers Management from Pune University (1999) and a BSc (Computer Science) from Delhi University (1997), India. She is pursuing her PhD in Computer Science in the area of data mining. She is currently employed as an Associate Professor with Amity University Uttar Pradesh, Noida, India and has 13 years of experience in IT field. Her research interests include data mining and its applications.

Ajay Rana is Director, Professor and Advisor as well as the Founder Director of various centre/institutions of Amity Universe. He obtained his PhD and MTech in computer science and engineering. He has published more than 81 research papers in reputed journals and proceedings of international and national conferences. He has co-authored four books and co-edited 15 conference proceedings. He has delivered invited lectures in more than 30 technical and management workshop/conferences programs in India and abroad. He has organised 45 plus conference, workshop, faculty development program.

Jayanthi Ranjan completed her PhD from Jamia Milia Islamia Central University, India in the area of data mining and has more than 15 years of experience. Her research contribution includes papers on data mining, data clustering, database security, business intelligence in various industry domains. She has publications in journals like, *International Journal of Business Information Systems*, *International Journal of Indian Culture and Business Management*, *Int. Journal of Business Innovation and Research*, *International Journal of Business Excellence*, *Information Management and Computer Security* and many more. More than 50 research publications are there in international journals and conferences. She is currently employed with IMT Ghaziabad, India as a Professor in Information Management and Systems, the Associate Dean in International Relations, Editor-in-Chief of *International Journal of Intercultural Information Management* and also editor of various others reputed international journals.

1 Introduction

Many researchers have done wide-ranging exertion in depicting and maximising the customer lifetime value (CLV) (Berry and Linoff, 2004; Chao and Yang, 2003; Kim and Kim, 1999; Kim et al., 2006; Kim and Street, 2004; Kumar and Reinartz, 2005a, 2005b) and to learn the strategies involved in retaining the customer. Customer relationship management (CRM) (Ranjan et al., 2008; Ranjan, 2010; Kumar, 2006; Vijayaraghavan and Kannan, 2011) details about customer acquisition, customer retention and even customer churn rate. Customer interaction and contribution is quite high in those companies which manufactures the product and provides after sale service.

Social interactions and knowledge of a customer to new products enhances competition in the market among the manufacturers. Various marketing induction programs are initiated to increase the stay of customers' time period with them. This stay of customers can be quantified in value using CLV model. CLV is a measure which computes total contribution of customer in company's cash flow over the period of time. The assessment of CLV summarises the future success of the organisation. CLV helps in identifying the value of potential customers (Gupta and Ames Stuart, 2004). CLV affects the market from customer acquisition to retention stage (Miglautsch, 2002, 2000; Vijayaraghavan and Kannan, 2011; Rust et al., 2004). CLV helps to know the worth of the customer to the management which in turn can build strategies to retain the loyal customer by ranking or segmenting process. CLV is a very relevant and well adopted exercise used in various big firms. Many companies like IBM, Amazon, ebay they use CLV as the tool for calculating the revenue of the company.

Maximisation of CLV is supported by many researchers. Diverse dimensions of CLV have been researched and applied by experts. Marketing induction programs and strategies are well explained in automobile industry by Yoo and Hanssens (2005). Reinartz et al. (2003) presented a new insight in customer value by analysing that not always the past information contributes in the future purchase of the customers. Many authors also felt the same and strongly emphasised on the definition of quantitative or objective metric in calculating CLV. Kumar and Reinartz (2005a) proposed share-of-wallet metric for a focal brand. Gupta and Lehmann (2003) computed average CLV and Kumar and Ramani (2004) customer purchase and evaluation in competitor's firm. Traditionally, used metrics (Gupta and Valarie, 2006; Kim et al., 2006) like RFM, share-of-wallet and past customer value (PCV) gives the measurement of computing CLV with the help of quantitative measures. But there are various other antecedents which affect the purchasing behaviour of the customers. Gupta and Valerie (2006) emphasises on the firm's influence on customer behaviour. Thomas (2001) proposed customer acquisition and retention are related together. CLV defines metrics which can be calculated as an individual and aggregate approach. Customer equity (CE) defined by Thomas as the summation of customer acquisition, retention and return. Average CLV is measured by dividing CLV with CE. Rust et al. (2004) importance of customer is quite vital as losing a customer is a big loss since customer does not return back to the suppliers. Kumar explained that further research in CLV has to be done by considering hidden factors influencing customer purchase behaviour. In this paper hidden factors that influences the customer lifetime period with the firm is presented. Lin and Tang (2006) shown in his paper customer clustering for a music company product using RFM model and then with Apriori algorithm created association rules to recommend music product for the potential customers. Chen and Zhu (2010) followed the similar approach using RFM variables and K-means algorithm to classify customers and further using rough set theory to generate rules for CRM system. Chao and Yang (2003) worked for medical equipment industry. K-means and SOM has been used for market segmentation by Kuo et al. (2006) in fright transport industry. Shin and Sohn (2004) used fuzzy K-means cluster in stock market for customer segmentation. LTV model is applied by Hwang et al. (2004), and Kim et al. (2006) in wireless telecommunication to develop strategy and customer segmentation. Many researchers improvised RFM model by evaluating other contributions made by the customers, thus adding customer value to the company.

In this paper, an empirical research analysis is done, by using proposed model based on CLV concept for the car manufacturing industry, markets clusters are found which are then mined by using one of the association rules algorithm (Chiang, 2008, 2010), direct hashing and pruning (DHP) (Park at al., 1995) and customer patterns are generated.

Data mining technique explores the hidden trends from the huge pool of data with one of its popular method association rule mining. The analysis of data generates useful patterns (Singh et al., 2012; Han and Jian, 2000) and rules for business purposes. Pattern mining facilitates decision makers to categorise and summarise marketing decisions based on the data analysis. Agarwal et al. (1993) propose association rule mining as market basket analysis concept. It defines the relationship between different itemsets from the database.

Objective of the paper: The current paper investigates customer purchasing capability in various dimensions. Significant and essential dimensions are considered which adds value to the active customer retention. A framework for the proposed model is explained using data mining approach and results are found by taking an example of automobile industry, hugely contributing in market share.

The paper is organised as follows: the predefined traditional metrics for determining customer value is discussed in Section 2. In Section 3, proposed customer centric analytical model is explained. In Section 4, it briefs about the data mining approach in PCRM model framework. In Section 5, an empirical research analysis is used to discover results. Section 6 implications of study and Section 7 cover the conclusion and future scope of work.

2 Pre-defined traditional metrics for determining customer value

Recency, frequency and monetary (RFM): these three metrics are used as a score model to find out the most likely purchase made by the customers in near future over period of time. RFM helps in categorising the customer value. Where R measures last when the customer purchased the product; F calculates how often a customer purchases the product over a period of time and M tells the money spent by the customer. RFM is a scoring model which results in the score of customers. RFM model is to separate each of variables into five levels. First level is the lowest level and the fifth level is the highest level (Gupta and Ames Stuart, 2004). Thus, R-F-M can be classified from 1-1-1 to 5-5-5.

Past value customer (PCV): it predicts total/gross contribution made by the customer in the future based on the past purchase value done by the customer over a period of time. It finds the cumulative contribution till the present period.

2.1 Limitations

The above metrics limits the scope in knowing the customers’ future value but do not reveal the factors that would lead the customers in the category of active customers and the behaviour, influence and mind set of customers’ which led the customers’ to purchase product/service from the specific organisations. Also, PCV metrics alone cannot depict the expected maintenance cost of the customer (Kumar). RFM only works with the historical customer data not with the forecast data. Since, RFM is a scoring model not suitable to measure dollar amount of CLV. The challenging task is to predict whether customer is going to purchase product/service in the future with the company and how company can maximise the return.

2.2 Benefits of CLV

CLV is an important measure to calculate as it gives insight in marketing strategy planning’s like,

a it involves identification of potential customers and planning to increase the degree of retention of customers

b to identify customer leaving the brand product (Reinartz et al., 2002)

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.

3 Proposed analytical customer centric model (PCRM)

For manufacturing (Harding et al., 2006) industry like, automobile the customers can immensely contribute in various ways. Today's, customer is very active on social websites. In today's era, one cannot deny the power of recommendation about product/service by customer on social websites like Twitter, Facebook, etc. In this model PCRM, the proposed model equation has an upper edge in calculating CLV over RFM and PCV metrics (McCarty and Hastak, 2007). In the proposed model PCRM, customer is termed as active customer because it has been observed from the data that customer might be associated with the organisation for many years but remain there in passive state means not recommending or referencing the product/service to anyone as well as not buying any other variant/vehicle from the organisation, or may be the customer was active in the initial years of purchase but after that for many years not responding to the organisational facility. In the proposed model for automobile sector authors emphasised that referencing/recommendations/review posts on social websites (Bonchi et al., 2011) enforces the purchasing behaviour (Linoff and Berry, 2002) of the new customers. Here, in this model the customer is assumed as an active customer who is actively participating and interacting directly or indirectly with the organisation.

3.1 Key drivers used in proposed metric (PCRM)

For developing a customer centric analytical model (Colombo and Jiang, 1999) the kind of information required is as follows:

a the demographic/vital data about customers like age, occupation, career , income, etc.

b customer purchase/buying cost of variant

c customer purchase details about accessories

d customer frequency of availing service from the dealer

e customer source of enquiry (recommendations by existing customers, review posted)

e miscellaneous relevant customer information like insurance, warranty period and exchange_of_old_car.

Using the above parameters customer centric analytical model can be dimensions can be designed as follows:

$$\text{Active Customers' Retention Value } \left(AC_R V\right) = P + C + R + M$$

P (at time *t*) = AP_t	direct purchase of variant made by the customer (directly adding to the company cost)
C (at time *t*) = AC_t	Costs in purchasing types of accessories
R (at time *t*) = AR_t	active customer review/opinion about on the social website
M (at time *t*)= AM_t	miscellaneous variable/costs involved in availing service/insurance/exchange_of_old MarutiCar/warranty.

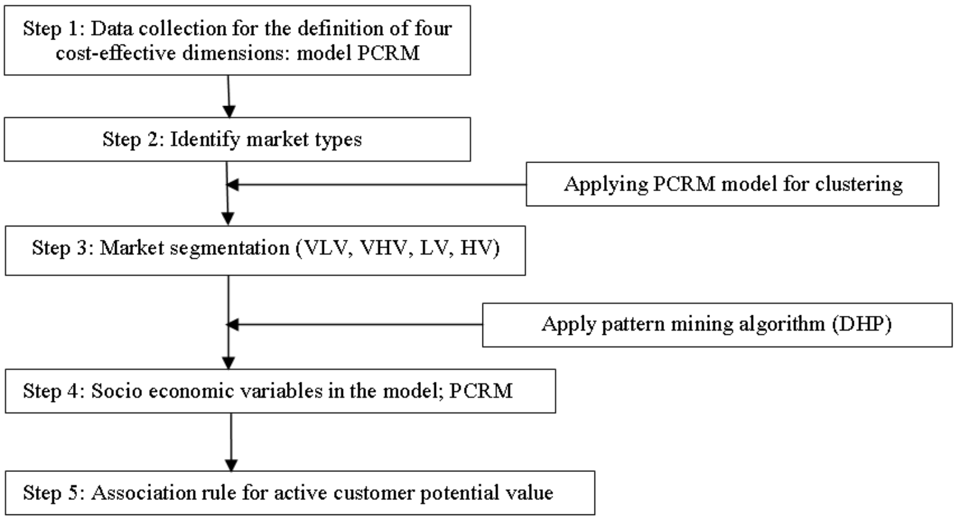
For manufacturing companies, the second parameter which contributes to customer retention value, i.e., AC_t (active customer current value) includes:

- Mathematical equation is given in Appendix.

4 Data mining approach

The approach based on data mining used in finding the observations for model PCRM is as follows:

Figure 1 Conceptual approach of research study



Using the ranking based method from 1 to 4 levels from the high valued to low valued (HV to LV). Data is fed according to the PCRM model variables and looking after customer contribution in the selling of product /service categorises the customer values into very high valued (VHV), very low valued (VLV), low value and high value. The generated Market clusters are analysed and socio economic variables of customer data using DHP algorithm to generate marketing patterns of customers.

The questionnaire of this research is mainly focusing in design of database structure for knowledge discovering. RFM model is not an applicable for every kind of business application. Therefore, PCRM model is proposed and applied to find out the business solution of an automobile industry. Lin and Liu (2008) explored the RFM variables for studying behaviour of online booking o air passengers.

The study *uses* questionnaire method to collect data. The questionnaire consists of two parts: variables of PCRM (RFM model-based), and socio-economic variables. All collected data are in two phases primary and secondary with the purposive sampling method. The sample size is measured by the formula: (Singh et al., 2011) $n = \frac{[Z^2 \times P(1-P)]}{e^2}$, where n is the sample size, confidence interval: 95% (Z: 1.96), e: 0.05, P-ratio: 0.5.

4.1 Association rule mining algorithm DHP (Park et al., 1995)

DHP algorithm enhances the performance of Apriori algorithm. The Apriori (Agarwal et al., 1993; Agarwal and Srikant, 1994; Berry and Linoff, 2004) algorithm explores frequent itemsets from the transaction data to find the most occurring patterns. This interesting association between items sets (Reinartz et al., 2003; Rust et al., 2004) knowledge benefits the organisation. The algorithm moves with creating a candidate set (i.e., a list of possible item combinations) and then counting the occurrences of each candidate itemsets in the transactions. The frequent ones (Zhou et al., 2007) whose count is equal or above a certain acceptable level (minimum support) are then selected into the frequent itemsets. The algorithm uses hashing function to assign bucket # to two-candidate itemset and bucket count provides the frequent itemsets. It uses a hash function as:

$$H(x, y) = \big((\text{order of } x * 10) + (\text{order of } y)\big) \bmod n, \text{ where } n \text{ is the arbitrary number.}$$

Here, x and y are itemsets in the database.

4.1.1 Steps of DHP algorithm flow to generate frequent itemsets from the database

Step 1 Input: clusters and socio-economic variables with their frequent occurrence (support count).

Step 2 Set minimum support count value to MinSupp variable.

Step 3 Generate frequent relation set (L_k) from candidate set using self join L_{k-1} with itself using hash function, $H(x, y)$.

Step 4 $H(x, y)$ = ((order of x * 10) + (order of y)) mod n, where n is the arbitrary number.

Step 5 Compute for any pair of candidate set (x, y) the bucket value with the hash function.

Step 6 Count all the candidate set in each bucket.

Step 7 Pruning: If bucket value of a candidate set is less than MinSupp then remove the candidate relation set from the hash table.

Step 8 Repeat the process until all frequent relations sets are identified.
Store the frequent relations sets.

Step 9 Stop.

4.1.2 Rule generation

A transaction t is said to contain A if and only if $A \subseteq t$. Now, mathematically, an *association rule* will be an implication of rule for $A \Rightarrow B$ where A and B both are subsets of Ω and $A \cap B = j$ (the intersection of sets A and B is an empty set).

The *support* of an itemset is the fraction the rows of the database that contain all the items in the itemset. *Support* indicates the frequencies of the occurring patterns. Sometimes it is called *frequency*. Support is simply a probability that a randomly chosen Transaction t contains both itemsets A and B. Mathematically:

$$Support\ (A \Rightarrow B)t = P\left(A \subseteq t \bigcup B \subseteq t\right)$$

Will be shown in a simple notation as:

$$Support\ (A \Rightarrow B) = P\left(A \bigcup B\right)$$

Confidence: It defines the strength of implication in the rule. Sometime it is called *accuracy*. Confidence is simply a probability that an itemset B is purchased in a randomly chosen transaction t given that the itemset A is purchased. Mathematically,

$$Confidence\ (A \Rightarrow B)t = P\left(B \subseteq t \middle| A \subseteq t\right)$$

Can be shown with a simple notation as:

$$\begin{aligned} Confidence\ (A \Rightarrow B)t &= P\left(B \middle| A\right) \\ &= P(B) / \left(A \bigcap B\right) \end{aligned}$$

f frequency of itemset appeared in the database

N number of transactions in the database.

Support can be defined as, $S = (f / N) * 100\%$.

$$\text{Confidence : If } A,\ B \ -> \ C \text{ then } Support\ \ (A, B, C) / Support(A, B) * 100\%,$$

Here, A and B are called antecedents and C is called consequent.

5 Empirical research analysis

The automotive industry in India is one of the sixth largest in the world (Thomas, 2001; Triola and Franklin, 1994). 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 the present paper, an example of an automobile industry in India is taken into research study. In India, the variant is manufactured keeping its customer and roads structure in mind. There is lot of automobile companies manufacturing variants. Maruti Suzuki better known as Maruti is one of the established brands in India. Maruti believed to design variants those provide better mileage, service and should fit in customers' budget.

5.1 Research design: data collection and sample size

The data collection is done in two phases: primary and secondary. The primary data is collected by various brainstorming sessions with the researchers attending International Conferences, symposiums and workshops, questionnaires sent via e-mails, face to face interaction with the customers (Goodman, 1992). The questionnaire were filled during the period March to September’2010.There were 351 questionnaires accurately, completely filled and 14 were invalid questionnaires. Thus, the rate was more than 96%.Followed the same computing formula as: $n = \frac{[Z^2 \times P(1-P)]}{e^2}$. The questionnaire is prepared based on PCRM variables and demographic details of the customers. Interaction and long discussions with the researchers led to accuracy of information. The secondary data was collected from the trusted Maruti dealer, contributing in size of 380 out of 631 totals. Sample size of 631 customers was preprocessed and analysed.

Step 1 Defining four cost-effective dimensions of PCRM model.

1. [P] Purchasing costs: Purchasing costs of the variant is determined by categorising the variant into categories luxury segment (4), high segment (3), middle segment (2) and average segment (1).
2. [C] Accessories cost: Costs involved in buying extra accessories with each category variant/vehicle. Additional accessory purchased with luxury variant 2 to 3 types of accessories (1), high segment variant accessories 3 to 4 (2), middle segment accessories 4 to 5 (3) and average segment accessories 5 or more (4).
3. [R] Recommendation and Review posts on the social website poll about the variant: Number of recommendations/ reviews posted by the customers on the social website. It is calculated with the source of enquiry attribute and by checking the website reviews .The number of times each variant is recommended by the customer. Frequency (4), (3), (2) and (1) are assigned.
4. [M] Miscellaneous variable involved: Frequency of old variant exchange (4), (3), (2) and (1) are assigned respectively from high to lowest.

The research study uses the questionnaire and the social website poll about each variant as well as the customer transaction data having source of enquiry as an attribute to know the recommendations done by existing customers.

Step 2 Identifying PCRM markets

According to Berry and Linoff (2004) and Chiang (2010), the profit increasing is the basis of customer value. Pattern generation is done using socio-economic variables ranging from VLV to VHV. The market is ranked into four levels from very high value to very low value (1-1-1-1 to 4-4-4-4). Each profitable variable of PCRM model can be seen as VHV and VLV .Thus, there would be more markets but in the figure below the relevant and essential markets are shown. Four-dimensions (PCRM) into four levels will generate 64 markets. Non-significant are pruned and significant and essential markets are further categorised into relevant four types.

Step 3 Market segmentation analysis (Cheng and Chen, 2009).

Step 4 Applying socio-economic variables like, gender, age, income, occupation, number of family members, number of Maruti variants into association rule mining algorithm (DHP) to generate frequent market patterns.

Step 5 Association rule for active customer potential value.

Figure 2 PCRM markets (see online version for colours)

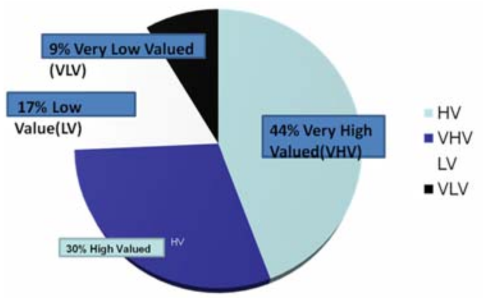
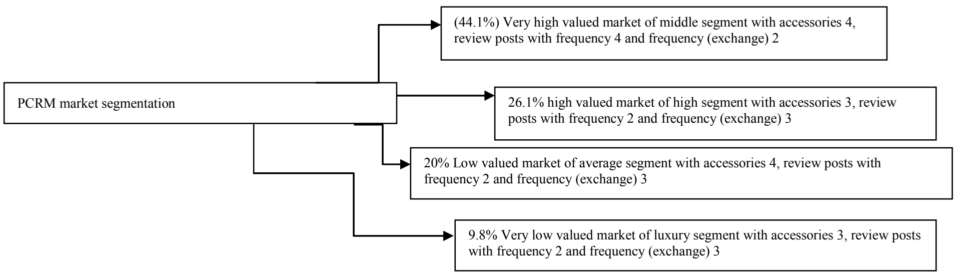


Figure 3 Segmentation of PCRM model market



5.1.1 Significant rules template generated can be classified as

$$< antecedent1,\ antecedent2,\ antecedent3\ldots >$$
$$=>< consequent1,\ consequent2,\ consequent3\ldots >$$

8,538 rules are generated with different values of consequent and antecedent, only significant rules are mentioned here.

- *Rule #1: support 98.20%*

  $<$ Gender ('Male') $\cap$ Age (21–30) $\cap$ Income ($>$ 4.5 to 6.5)
  $\cap$ Occupation ('Employed') $\cap$ Number of family members (2 to 3)
  $\cap$ Number of Maruti Cars('1') $>$ => P(VHV) $\cap$ C(HV) $\cap$ R(VHV) $\cap$ M(VLV)

- *Rule #2: support 90%*

  $<$ Gender ('Male') $\cap$ Age (31–40) $\cap$ Income ($>$ 12.5 to 16.5)
  $\cap$ Occupation ('Self-employed') $\cap$ Number of family members (5 to 6)
  $\cap$ Number of Maruti cars ('2') $>$ => P(VHV) $\cap$ C(LV) $\cap$ R(VHV) $\cap$ M(LV)

- *Rule #3: support 38.7%*

  $<$ Gender ('Female') $\cap$ Age (31–40) $\cap$ Income ($>$ 8.5 to 12.5)
  $\cap$ Occupation ('Self-employed') $\cap$ Number of family members (5 to 6)
  $\cap$ Number of Maruti cars ('2') $>$ => P(VHV) $\cap$ C(LV) $\cap$ R(VLV) $\cap$ M(VHV)

- *Rule #4: support 38.2%*

  $<$ Gender ('Male') $\cap$ Age (41–50) $\cap$ Income ($>$16.5 to 20.5)
  $\cap$Occupation ('Employed') $\cap$ Number of family members (5 only)
  $\cap$Number of Maruti cars ('2') $>$ => P(HV) $\cap$ C(VLV) $\cap$ R(VLV) $\cap$ M(HV)

- *Rule #5: support 37%*

  $<$ Gender ('Female') $\cap$ Age (41–50) $\cap$ Income ($>$12.5 to 16.5)
  $\cap$Occupation ('Self-employed') $\cap$ Number of family members (5 to 6)
  $\cap$Number of Maruti cars ('2') $>$ => P(HV) $\cap$ C(HV) $\cap$ R(LV) $\cap$ M(HV)

- *Rule #6: support 98%*

  $<$ Gender ('Female') $\cap$ Age (41–50) $\cap$ Income ($>$12.5 to 16.5)
  $\cap$Occupation ('Employed') $\cap$ Number of family members (4 only)
  $\cap$Number of Maruti cars ('2') $>$ => P(HV) $\cap$ C(LV) $\cap$ R(VHV) $\cap$ M(VHV)

Table 1 Pattern generation for Maruti variant customers

Rule #	*Gender*	*Age*	*Income (lac per annum)*	*Occupation*	*Number of family members*	*Number of Maruti cars*	*Pattern (PCRM market)*	*Support*
1	Male	21–30	> 4.5 to 6.5	Employed	2 to 3	1	HV-HV-VHV-VLV	98.20%
2	Male	31–40	> 12.5 to 16.5	Self-employed	5 to 6	2	VHV-LV-VHV-LV	90%
3	Female	31–40	> 8.5 to 12.5	Self-employed	5 to 6	2	VHV-LV-VLV-VHV	38.7%
4	Male	41–50	> 16.5 to 20.5	Employed	4 only	2	HV-VLV-VLV-HV	38.2%
5	Female	41–50	> 16.5to 20.5	Self-employed	5 to 6	2	HV-HV-LV-HV	37%
6	Female	41–50	> 12.5 to 16.5	Employed	4 only	2	VHV-LV-VHV-VHV	98%

6 Implications of the study

It has been observed that the customers belonging to market HV-HV-VHV-VLV (Gender: Male, Age = 21–30, Income > 4.5 to 6.5 lac per annum, Occupation: Employed, Number of family members: 2 to 3 Number of Maruti cars = 1). The market pattern of rule #2 is self employed males Age = 31–40, Income: >12.5 to 16.5, 2 Maruti product cars, with number of family members 5 to 6. Similarly, the pattern of rule #6 are females (Age = 41–50, Income: > 12.5 to 16.5, Occupation: Employed, Number of family members: 4 and Number of Maruti cars: 2). These customers are posting positive reviews on the website and their recommendation value of the product is again high.

The rule #3, 4 and 5 are the customers who are contributing less to PCRM model value but they are high in exchange of cars and low in posting positive reviews about the product/service on the social website or the recommendation value about the product to other customers is very low. From the above results, it is apparent that in PCRM Model the third dimension contributes the most, which is the number of recommendations done by the customer socially. In rule #3, 4 and 5 the direct variant cost has the high value, accessories cost is also high but since the recommendation value about the product is low, which is directly putting impact on the fourth dimension where frequency of exchanging old car is quite high.

The organisation can build up promotion policy for the existing customers to offer some discount on the number of times they recommend the product to other customer's. Also, the positive comments on the social website should also carry some monetary value points. This can convert the passive customers into active potential customers and can maximise their lifetime value with the organisation.

Rule #1, 2 and 6 focuses in identifying strong potential valued customers. Organisation can offer free membership for availing service facility and accessories. Rule #2 and 6, suggests that families having two Maruti cars adding high value to the direct variant purchase cost but low value in purchasing accessories. The organisation can build up intelligent marketing strategy for converting accessories low value (LV) to high value (HV). The study helps the organisation to upgrade their Marketing policies and CRM system according to the proposed model PCRM.

6.1 Business value to the organisation

The above results obtained can provide market strategy to the organisation.

Customer life cycle	*Objective*	*Interpretation of rules generated*	*Market strategy*	*Campaign event*
Acquisition	To attract new potential customers	Rule #1, factor [R] posting positive reviews about the high segment car [P]	Program for target customers of age 31–40 male income group 4.5 to 6.5 lac	Telecalling, free gift, family free drive (trial) of car
Retention	Retain customers possible to leave	Rule #3, 4 and 5, factor 4th [M] exchange of old car is high, [P] customers of luxury car with more number of accessories [C]	Strategy to introduce new accessories in luxury cars, target customers of age 31–50, male, income group 8.5–20.5	Free petrol cards, extend warranty period, lucrative offer with exchange_of_old_car
Potential	To increase more values with profitable customers	Rule #2 and 6, customers adding high value to the direct variant purchase cost [P] but low value in purchasing accessories [C]	To build intelligent policy to increase the conversion rate of low valued accessories to high valued accessories. target customers of Age around 40 years, income group 12.5–16.5	Free fuel filling card, free annual service facility

7 Conclusions and future scope of the work

It is a proven fact that data analysis using data mining techniques helps the organisation for better decision making and marketing strategy planning. For the growth of any product and service oriented industry, to keep bringing new policies to attract new and retain existing customers is a big and never ending challenge. In this research study, a new customer centric analytical model PCRM has been proposed for maximising lifetime value of the customers. This paper discovers the types of market and patterns with an example of an automobile industry, applying PCRM model and data mining technique association rule mining (DHP) for maximising CLV. The results show that keeping PCRM factors helps the organisation to be more strategic and competitive leader in the market.

Further, the work can be enhanced by exploring PCRM market for large transactional databases. This study further can also be improvised by comparing between various classification or association rules mining algorithm like decision tree or FP tree. Advantages and disadvantages of these techniques can be investigated. Secondly, more socio-economic parameters can be used for optimal breakthrough. Empirical data analysis and comparative results from other existing models and PCRM would give better insight to give the organisation about the factors influencing customer purchasing behaviour.

References

Agarwal, R. and Srikant, R. (1994) 'Fast Algorithms for mining association rules in large databases', *Proc. of 20th Int. Conference on Very Large databases*, Santiago de Chile, pp.487–489.

Agarwal, R., Imielinski, T. and Swami, A. (1993) 'Mining associations between sets of items in large databases', *Proc. Of ACM SIGMOD Int. Conference on Management of Data*, Washington, DC, pp.207–216.

Berry, M.J.A. and Linoff, G.S. (2004) *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*, 2nd ed., John Wiley and Sons, New York, NY.

Bonchi, F., Castillo, C., Gionis, A. and Jaimes, A. (2011) 'Social network analysis and mining for business applications', *ACM Transactions on Intelligent Systems and Technology*, Vol. 2, No. 3, Article 22.

Chao, C.M. and Yang, C.T. (2003) 'Applying data mining on prediction of customer value', *Conference of an Approach to Analyze and Implement Commercial Database*, Taipei, Taiwan, pp.159–183.

Chen, Y-M. and Zhu, M-L. (2010) 'Sampling based N-hash algorithm for searching frequent itemset', Published in the *Proceedings of Internet Technology and Applications (IEEE) Conference*.

Cheng, C.H. and Chen, Y.S. (2009) 'Classifying the segmentation of customer value via RFM model and RS theory', *Expert Systems with Applications*, Vol. 36, No. 3, pp.4176–4184.

Chiang, W.Y. (2008) 'Establishment of electronic marketing model for internet shopping: a case study of clothing and cosmetic categories', *Journal for SME Development*, Vol. 11, No. 2, pp.169–196.

Chiang, W.Y. (2010) 'To mine association rules of customer values via a data mining procedure with improved model: an empirical case study', *Expert Systems with Applications*, Vol. 38, No. 3, pp.1716–1722.

Colombo, R. and Jiang, W. (1999) 'A stochastic RFM model', *Journal of Interactive Marketing*, Summer, Vol. 13, No. 3, pp.2–12.

Goodman, J. (1992) 'Leveraging the customer database to your competitive advantage', *Direct Marketing*, Vol. 55, No. 8, pp.26–27.

Gupta, S. and Ames Stuart, J. (2004) 'Valuing customers', *Journal of Marketing Research*, Vol. 41, No. 1, pp.7–18.

Gupta, S. and Lehmann, D.R. (2003) 'Customers as assets', *Journal of Interactive Marketing*, Vol. 17, No. 1, pp.9–24.

Gupta, S. and Valarie, Z. (2006) *Customer Metrics and Their Impact on Financial Performance*, Working paper, Columbia University, New York.

Han, J. and Jian, P. (2000) 'Mining frequent patterns by pattern-growth: methodology and implications', *ACM SIGKDD*, December, Vol. 2, No. 2, p.31.

Harding, J.A., Shahbaz, M., Srinivas and Kusaik, A. (2006) 'Data mining in manufacturing: a review', *Journal of Manufacturing Science and Engineering, ASME*, November, Vols. 128/969, No. 4.

Hoekstra, J.C. and Huizingh, E.K.R.E. (1999) 'The lifetime value concept in customer-based marketing', *Journal of Market Focused Management*, Vol. 3, No. 3, pp.257–274.

Hwang, H., Jung, T. and Suh, E. (2004) 'An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry', *Expert Systems with Applications*, Vol. 26, No. 1, pp.181–188.

Kim, B.D. and Kim, S-O. (1999) 'Measuring up selling potential for life insurance customers', *Journal of Interactive Marketing*, Vol. 13, No. 4, pp.2–9.

Kim, S-Y. and Jung, T-S. et al. (2006) 'Customer segmentation & strategy development based on customer lifetime value: a case study', *Expert Systems with Applications*, Vol. 31, No. 1, pp.101–107.

Kim, Y.S. and Street, W.N. (2004) 'An intelligent system for customer targeting: a data mining approach', *Decision Support Systems*, Vol. 37, No. 2, pp.215–227.

Kumar, V. (2006) 'Customer lifetime value', in Grover, R. and Vriens, M. (Eds.): *Handbook of Marketing Research*, pp.602–627, Sage Publications, Thousand Oaks, CA.

Kumar, V. and Ramani, G. (2004) 'Taking customer lifetime value analysis to the next level', *Journal of Integrate Communications*, Vol. 4, No. 1, pp.27–33.

Kumar, V. and Reinartz, W.J. (2005a) *Customer Relationship Management: A Data Based Approach*, Wiley, New York.

Kumar, V.R. and Reinartz, W. (2005b) *A Purchase Sequence Analysis Framework for Targeting Products, Customers and Time Period*, University of Connecticut.

Kuo, R.J., Ana, Y.L., Wanga, H.S. and Chung, W.J. (2006) 'Integration of self organizing feature maps neural network and genetic K-means algorithm for market segmentation', *Expert systems with Applications*, Vol. 30, pp.313–324.

Lin, C.S. and Tang, Y.Q. (2006) 'Application of incremental mining and customer's value analysis to collaborative music recommendations', *Journal of Information, Technology and Society*, Vol. 6, No. 1, pp.1–26.

Lin, H.S. and Liu, I.H. (2008) 'Application of data mining to explore international airline passenger online booking behavior', *Transportation Planning Journal*, Vol. 37, No. 2, pp.197–236.

Linoff, G.S. and Berry, M.J.A. (2002) *Mining the Web: Transforming Customer Data into Customer Value*, John Wiley and Sons, New York, NY.

McCarty, J.A. and Hastak, M. (2007) 'Segmentation approaches in data-mining: a comparison of RFM, CHAID, and logistic regression', *Journal of Business Research*, Vol. 60, No. 6, pp.656–662.

Miglautsch, J.R. (2000) 'Thoughts on RFM scoring', *Journal of Database Marketing*, Vol. 8, No. 1, pp.2–7.

Miglautsch, J.R. (2002) 'Application of RFM principles: what to do with 1-1-1 customers?', *Journal of Database Marketing*, Vol. 9, No. 4, pp.319–324.

Park, J.S., Chen, M.S. and Yu, P.S. (1995) 'An effective hash based algorithm for mining association rules', *1995 ACM SIGMOD Int'l Conf. Management of Data*, May, pp.175–186.

Ranjan, J. (2010) 'Integrating analytical customer relationship management with customer relationship management in organizations', *International Journal of Innovation and Learning. IJIL*, Vol. 8, No. 4, pp.408–426, Inderscience Publishers.

Ranjan, J. et al (2008) 'Distinguishing operational customer relationship management and analytic customer relationship management – an organizational perspective', *International Journal of Logistics Economics and Globalisation*, Vol. 1, Nos. 3/4, pp.307–329.

Reinartz, W. et al. (2002) 'The mismanagement of customer loyalty', *Harvard Business Review*, Vol. 80, No. 7, pp.86–94.

Reinartz, W. et al. (2003) 'The impact of customer relationship characteristics on profitable lifetime duration', *Journal of Marketing*, Vol. 67, No. 1, pp.77–99.

Rust, R.T., Lemon, K. and Zeithaml, V. (2004) 'Return on marketing: using customer equity to focus marketing strategy', *Journal of Marketing*, Vol. 68, No. 1, pp.109–126.

Shin, H.W. and Sohn, S.Y. (2004) 'Segmentation of stock trading customers according to potential value', *Expert Systems with Applications*, Vol. 27, No. 1, pp.27–33.

Singh, A. and Ranjan, J. 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.

Singh, A., Chaudhary, M., Rana, A. and Dubey, G. (2011) 'Online mining of data to generate association rule mining in large databases', Published in the *Proceedings of International Conference on Recent Trends in Information Systems (ReTIS'2011)*, IEEE.

Singh, A., Rana, A. and Ranjan, J. (2012) 'Enterprise mobility in higher education universities: data mining perspective', *Proceedings of International Conference FOBE 2012*, 29–30 November, Bloomsbury Publishers.

Thomas, J. (2001) 'A methodology for linking customer acquisition to customer retention', *Journal of Marketing Research*, Vol. 38, No. 2, pp.262–268.

Triola, M.F. and Franklin, L.A. (1994) *Business Statistics: Understanding Populations and Processes*, Addison Wesley Publishing, NY.

Vijayaraghavan, R. and Kannan, P.V. (2011) 'Applications of data mining and machine learning in online customer care', *17th ACM SIGKDD International Conference.*

Wu, B., Ye, Q., Yang, S. and Group CRM (2009) 'A new telecom CRM framework from social network perspective', *ACM First International Workshop on Complex Networks Meet Information & Knowledge Management*.

Yoo, S. and Hanssens, D.M. (2005) *Modeling the Sales and Customer Equity Effects of the Marketing Mix*, Working paper, Anderson School of Management. University of California, Los Angeles.

Zhou, Q., Chen, Y. and Wu, H. (2007) 'A new algorithm finding frequent itemsets based on minimum separating into buckets', *Workshop on Intelligent Information Technology Application*, 2–3 December, pp.38–42, IEEE.

Websites

http://en.wikipedia.org/wiki/Automotive_industry_in_India.

http://en.wikipedia.org/wiki/Maruti_Suzuki.

Appendix

TC_t Customer total contribution which is dependent on parameter AC_t, AR_t.

$$TC_t = \Sigma\ (N \times Q) + L$$

where

N category of variants

Q price of each category of variants

L other relevant integrated income

AC_t customer current cost contribution.

$$AC_t = \Sigma(N \times U + R + A \times N)$$

where

U cost per category of variant

R risk cost associated with variant

A cost of accessories purchased with each category variant.

$$AR_t = MR + NR$$

MR customer reviews post on the social websites

NR customer purchase response from online campaign of the organisation.

$$\begin{aligned}&\text{Active customer prospective value of automobile sector}\\&= \text{Customer purchase value}\\&\quad + \text{Customer costsin purchasing in accessories}\\&\quad + \text{Active customer reviews about product / Service on the social website}\\&\quad + \text{Miscellaneous costs involved in availing service}\end{aligned}$$

So, customers' lifetime value (Hoekstra and Huizingh, 1999) is as follows for manufacturing sectors.

$$\begin{aligned}AC_R V_t = LTV_k = \sum_{t=0}^{p} ACS_{kt}(1+r)p-1 \\ + \sum_{t=p+1}^{n} CCH_{kt}\left(PS_{kt}\right)(1+r)p-t \text{ where, } t = 0\ldots p\ldots n\end{aligned}$$

LTV_k Customer's retention value when the time is at 'p' point where k is the starting time from the first transaction.

$$ACS_{kt} = f\left(r(s),(v),q\right)$$

$r(s)$ (sales of variant per time unit)

$p(v)$ (profit value contribution)

q (quantity of different variants).

$$CCH_{kt} = f\left(r(CR), o(k)\right)$$

CCH_{kt} customer hidden contribution during time period *kt*

$r(CR)$ *r*(customer's reference potential)

$o(k)$ (review post on social website)

PS_{kt} Satisfied customer = *f* (predicted sale figure, predicted profit)

r rate of discount.

Active Customer Retention Rate $\left(AC_R R\right) = 1 - CCR$ where,

CCR Customer churn rate, i.e., the % of customers who end their trust relationship with organisation in a given period of time.