2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)
Amity University, Noida, India. June 4-5, 2020

Customer Segmentation Using Fuzzy-AHP and RFM Model

Anu Gupta Aggarwal
Department of Operational Research,
University of Delhi
New Delhi, Delhi
anuagg17@gmail.com

Sweta Yadav
Department of Operational Research,
University of Delhi
New Delhi, Delhi
swetayadav.du.or.18@gmail.com

***Abstract*- In today's business environment the number of customers buying products online have increased, it is difficult for the firm to determine customer lifetime value (CLV) of every customer. In this paper, customer segmentation is used to find the customer lifetime value for a UK based registered company that sells unique gifts. Firstly, customer segmentation is done based on three characteristics of purchase variables: "Recency", "Frequency", and "Monetary". Next, weight or priority is given to the variables, which are attained using Fuzzy AHP technique. Lastly, analysis of the results is done by dividing the dataset into eights clusters and ranking those clusters on the basis of CLV values which helps the marketers in making selling and marketing decisions use these CLV values. Decision maker of the company consider technique of this paper in the future for calculating customers lifetime value (CLV) and gives more attention of that group which have highest CLV value.**

Keywords- Customer relationship management; Customer lifetime value; RFM analysis; Fuzzy-AHP; Customer segmentation.

I. INTRODUCTION

Due to high competition and complexity associated with the online business, it is difficult to know the actual contribution of a customer to the firm. Retaining of customers and gaining their loyalty has also become a tedious task. For establishing the appropriate customer relationship management (CRM) every company wants to recognize the customer's true value and loyalty by the data received from the customer transactions. Using CRM the companies make long-term relationships with customers which are valuable for the company. The concept of customer lifetime value (CLV) is more widespread than the CRM, and it is introduced in 1974 by Kotlar. Customer lifetime value defines the as the recent value of the customers purchase over the life. Calculation of CLV for individual customers is difficult then segmenting the customers. Loyalty (Chang, 2011) of customers are determine by the past purchasing behavior of that customers. RFM is one of the Customer lifetime value model that used for customer loyalty in lot of research models [1] . RFM be a behavioral model and all the past customer purchases to prospect the future customer behaviors.

In this paper, customer segmentation is applied on the transactional dataset of the UK based company that sells unique gift items. For this research, data was taken from the UCI machine-learning repository containing actual transactions. Analysis was done in two parts, where data selection, preprocessing, parameter extraction and clustering of the customers is done in the first step. In the second part Cheng extent analysis of Fuzzy-AHP method is applied for calculating the weights of the RFM parameters [2] followed by multiplication of normalized RFM values with the corresponding weights.

The rest of the paper is organized as follows: Section 2 discusses the theoretical background related to the study. The research framework and numerical illustration along with the results are discussed in Section 3. In Section 4, proofs of the validations of the results extracted from the study are discussed. In the last section, we talk about the conclusions drawn from the study.

II. THEORTICAL BACKGROUND

Customer Relationship Management (CRM) is a process in the organizational front, which aims at client acquirement and retention intentions. This is done by delivering what the customer wants and thus increasing their loyalties. It comprise of four dimensions such as identification, attraction, retention and development of customer base [3]. CRM challenges the product centric approach and emphasizes more on customer centric approach by empowering the customer and customer value creation [4]. To attain the aim of CRM, assessment of customer value is an important step in marketing domain [5]. Customer Lifetime Value (CLV) can be defined as the total value customer brings to the firm after the deduction of cost incurred in serving the customer, also taking into account the time value of money [6]. This value assists the strategists in classification of customers into various groups which further helps in designing appropriate strategies for different customer groups with different preferences [7] .

Majority of firms in the market have realized that dividing the customer base into segments is advantageous as compared to considering market as a whole. This helps in serving the diverse customer needs appropriately [8]. Customer segmentation has multiple benefits like increasing customer satisfaction along with increased revenues of the firm. It also assists in maintaining a good long term relationship with the customer [9]. Clustering is one of the most effective segmentation methods. In clustering, the whole group of data is divided into a number of homogeneous subgroups such that the intra group similarity is higher than the inter group similarity [10]. In this study, K- mean clustering has been utilized for the segmentation of data wherein the task is to simply classify the data into k groups on the basis of attributes, where k is a positive integer. For the grouping purpose, a centroid is identified and the sum of square of distance between the data points and centroid is minimized [11].

Fuzzy-AHP was proposed by [12], it is the combination of fuzzy set theory with the AHP technique. It captures both

978-1-7281-7016-9/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: University of New South Wales. Downloaded on October 24,2020 at 12:30:42 UTC from IEEE Xplore. Restrictions apply.

the logical thinking along with relative importance of the evaluation criteria. The main assumption of this technique is all the variables are independent from each other. This Fuzzy-AHP matrix of pairwise comparison comprises of triangular fuzzy conversion scale [13], which is then converted into crisp values. For doing conversion of the triangular scale into crisp follow the steps of the Cheng's extent analysis.

TABLE I. TRIANGULAR FUZZY CONVERSION SCALE [13]

Linguistic Scale	Triangular Fuzzy reciprocal	Triangular Fuzzy Scale
Absolutely more important	(2/7 , 1/3 , 2/5)	(5/2 , 3, 7/2)
Very Strongly important	(1/3 , 2/5, 1/2)	(2 , 5/2 , 3)
Strongly more important	(2/5 , 1/2 , 2/3)	(3/2 , 2 , 5/2)
Weakly important	(1/2 , 2/3 ,1)	(1 , 3/2 , 2)
Equally important	(2/3 ,1 ,3/2)	(1/2 , 2/3 ,1)
Just Equal	(1 ,1 , 1)	(1 ,1 , 1)

III. NUMERICAL ILLUSTRATION

The proposed framework of this paper is shown in **Fig. 1**

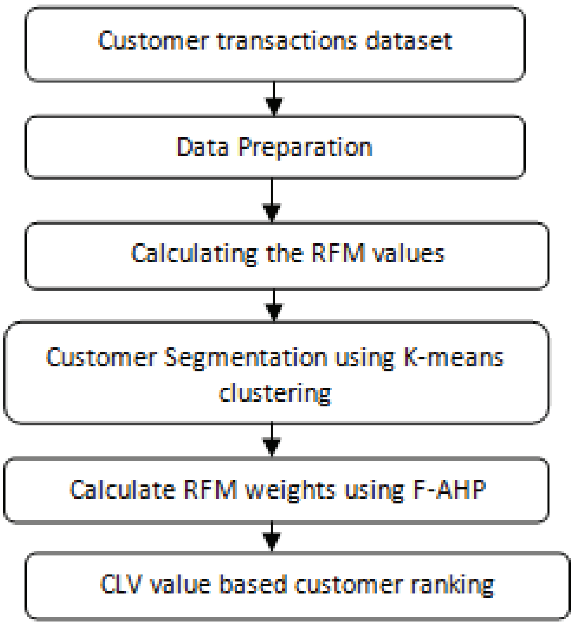

Fig. 1. Proposed framework of the paper

The whole process of this study is divided into six steps, which are shown in **Figure 1**. The Description of the steps are given below:

Step 1: Data extraction & preprocessing: We have used the transactional dataset of the UK based registered online retail firm available free at the Kaggle.com. This data contains 8 attributes including customer ID, Product invoice number, transactions date, unit price of product, description of product, stock number. Corresponding to each customer there are multiple transactions in a single day. Calculating the RFM values with respect to each customers.

TABLE II. DEFINITION OF PARAMETERS

Parameter	Definitions
Recency (R)	Interval between the time of the last transaction and the specified time period of the data extraction.
Frequency(F)	Number of times the transaction taken place by the each customer in the specified time duration.
Monetary(M)	Total amount of money spend by the each customer during the specified time interval.

We extracted the 590 customers transactions records after removing the cancelled order transactions. This dataset shown in the **Table 3.**

TABLE III. CUSTOMER TRANSACTIONS MATRIX

Serial No.	Customer ID	Recency (Days)	Frequency (number)	Monetary(pound)
1	12346	318	2	154367.2
2	12347	23	7	1545005056
3	12348	68	4	7021.92
4	12349	11	1	1757.25
5	12350	303	1	334.4
6	12352	29	10	154514.1
…	…	…	…	…
590	13115	29	4	8822.52

Step 2: For getting the optimal number of clusters to customer segmentation used the DUNN-INDEX because it is mostly used technique. In this method, maximize the inter-cluster distances and minimizes the intra-cluster distances. The Dunn-index, D for the partitions P.

$$D(P) = \frac{min_{i,j=1,2,3.,i\neq j}\delta(P_i, P_j)}{max_{i=1,2,..,k}\Delta(P_i)} \qquad (1)$$

where $\delta(P_i, P_j)$ denotes the distance between the clusters P_i, P_j (inter-cluster distance), and $\Delta(P_i)$ represents the distance Dunn index [14] between intra-cluster P_i, and k be the optimal number of clusters. DUNN-index algorithms given 8 optimum number of clusters.**Step 3:** The clustering of the customers into the eight groups because the 8 be the optimal number of cluster an that clustering was done by using SPSS software are shown in the **Table 4**.

TABLE IV. RESULT OF THE CLUSTERING ON THE BASIS OF RFM PARAMETER

Cluster	Result of the clustering			Pattern	Customer Segmentation
	R	F	M		
1	88.33	60.00	859539.04	R↑F↑M↑	Potentially unstable
2	23.00	7.00	1545005056	R↓F↓M↑	More Valuable
3	48.98	12.49	75003.32	R↑F↓M↓	Valueless
4	45.00	23.00	2845685.35	R↓F↑M↑	Valuable
5	35.50	25.50	561416.59	R↓F↑M↓	Less Valuable
6	34.89	20.56	265115.42	R↓F↓M↓	Valueless
7	110.29	3.39	4864.24	R↑F↓M↓	Valueless
8	21.00	194.00	5639987.40	R↓F↑M↑	More valuable

Authorized licensed use limited to: University of New South Wales. Downloaded on October 24,2020 at 12:30:42 UTC from IEEE Xplore. Restrictions apply.

According to **Table 4** cluster 2 and 8 customers are more valuable for firm, both of having high monetary value than the other cluster values. This group customers consider as loyal customers who spend a lot money on purchase and frequently purchased. Customers in cluster 3, 6, and 7 are valueless because of low frequency and monetary value also**.** Customers in clusters 1 having the special with having all three parameters values are high, are those customers who are occasionally buy or not loyal for one firm but because of high frequency and monetary value firms do not ignores this kind of customers, if the firm ignores this group they loose potential. Study of these customers called customer potential crisis customers.

Step 4: Normalizing the **table 4** of RFM parameter calculating the CLV value. Equ.2 for Normalized frequency and monetary value, for normalizing recency equation 3 used. This formula for normalizing [15] the data are given below:

$$x' = \frac{x - min_v}{max_v - min_v}(newmax_v - newmax_v) + newmin_v \quad (2)$$

$$x' = \frac{max_v - x}{max_v - min_v}(newmax_v - newmin_v) + newmin_v \quad (3)$$

Where min_v and max_v denotes the minimum and maximum values of the attribute v, respectively. Then $max_v - min_v$ normalization map a value x of v, to x' in the range of $[newmin_v, newmax_v]$.

TABLE V. NORMALIZING RFM VALUES MATRIX

Cluster	Normalized value of RFM		
	N_R	N_F	N_M
1	0.246	0.297	0.00055
2	0.978	0.0189	1
3	0.6866	0.0477	0.00046
4	0.7312	0.103	0.0184
5	0.838	0.116	0.0036
6	0.844	0.0901	0.00017
7	0	0	0
8	1	1	0.0036

Step 5: Calculate the CLV score for each cluster based on the RFM weights, as the formula in equation:

$$CLV = N_R \times W_R + N_F \times W_F + N_M \times W_M \quad (4)$$

Where N_R, N_F and N_M be the normalized values of Recency, Frequency and monetary respectively. The weights are denoted by W_R, W_F and W_M be the recency, frequency and monetary values.

These weights calculating by the Fuzzy-AHP technique because it remove the uncertainty of human judgement. The relative weights of the RFM variable are as follows: $W_R = 0.076$, $W_F = 0.321$ and $W_M = 0.603$. In the **table 6** contains the values getting by the multiplication of weights and normalized values with corresponding parameters.

TABLE VI. CLV RANKING ON THE BASIS OF CUSTOMER SEGMENTATION

Cluster	$N_R \times W_R$	$N_F \times W_F$	$N_M \times W_M$	CLV value	CLV rank
1	0.0187	0.0927	0.000332	0.1117	CLV3
2	0.0743	0.0061	0.603	0.6834	CLV1
3	0.0522	0.0149	0.000277	0.0674	CLV7
4	0.0556	0.0331	0.0112	0.0999	CLV5
5	0.0637	0.0372	0.0022	0.1031	CLV4
6	0.0641	0.0289	0.000102	0.093102	CLV6
7	0	0	0	0	CLV8
8	0.076	0.321	0.0022	0.3992	CLV2

IV. RESULTS VALIDATION

It is necessary to check the homogeneity between the within clusters of customers (membership degree), clusters having homogeneous (similar characteristics) are belong to same clusters otherwise different clusters. To analysis the membership degrees using the analysis of variance (ANOVA) [16] but it has one assumption for validity apply the Levene's test, because it based on the power and robustness. The null and the alternatives hypotheses for ANOVA analysis [17] as follow:

$$H_O: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 \ldots. = \sigma_8^2$$

$$H_1 = \sigma_i^2 \neq \sigma_j^2 for\ one\ i \neq j \quad (6)$$

Accept the alternate hypothesis no further verification is required for our results, all the clusters are significantly different and valid.

Solve this test by the SPSS at the 5% level of significance and the results of this test are shown in table. According to the table p value is 0.058 which is greater than the 0.05 so accept the null hypothesis, it conclude that the variance of different group are equal and assumption is true then proceed to ANOVA. In the Table 7 df1 and df2 represents degrees of freedom of clusters and the customer transactions respectively.

TABLE VII. LEVENE'S TEST RESULT

Variable	f-value	df1	df2	Significant level	Result
Membership degree of groups	2.361	4	582	0.058	Accept the null hypothesis

After apply the Levene's test for confirms the equality of the variance of clusters of different groups. Apply the ANOVA for determine whether there are any statistically significant differences between the means of two or more independent the group. This test taking place into the 1%, 5% or 10% level of significance but here the 5% level of significance have consider. The hypothesis are as follows for ANOVA at the 5% level of significance:

$$H_0: \mu_1 = \mu_2 = \mu_3 \ldots \mu_8 \quad (7)$$

$$H_1: \mu_i \neq \mu_j\ for\ one\ i \neq j$$

Run the ANOVA on the SPSS software for the above hypothesis. In the Table 8 result of the SPSS for ANOVA technique are given:

Authorized licensed use limited to: University of New South Wales. Downloaded on October 24,2020 at 12:30:42 UTC from IEEE Xplore. Restrictions apply.

TABLE VIII. RESULT OF ANOVA

	Sum of squares	Degrees of freedom	Mean square	f-value	significance
Between groups	85783434202	7	12254776315	48.476	0.00
Within groups	1.471 E+11	582	252800812.9		
Total	2.329 E+11	589			

According to the results of the ANOVA table, significance level is less than the 0.05 it concludes that the there is significance difference between means.

V. CONCLUSION

Customer segmentation is very important for CLV applications in many studies. Today online shopping is gaining popularity that's why this study is beneficial. For all this kind of online stores. Based on the results of this study, firms can formulate new as well as manage the old strategies for selling the products that are customers oriented. This study shows that monetary values are far more important than the other two parameters (Recency, Frequency).According to the our study Monetary (0.603), follow Frequency (0.321) and the least Recency (0.076). On the basis of these weights, ranked the customer segmentations. At the end, the table of ANOVA or validation table showed that, all the clusters are significantly different at 5% level of significance. A customer providing monetary benefits to the company is preferred over a customer who purchases more often or the one who has recently made a purchase. Each firm enjoys the liberty of choosing the preference of R, F and M values.

On the basis of this research, the company can making the new strategies, which are customer-oriented for increase the customer loyalty and maximizing the customer lifetime value. The firm focus more on the high CLV valued customer because they are high profitable, long time loyal for that firm. Company treated the customer differently on the basis CLV values because every company limited capability, so the managers of firms wants to maximize benefits within that resources. Decision-maker of the firm take the selling decision by the using the this paper methodology give the more attention to that customers who are more valuable, and this value getting by the CLV values of the clusters.

In this research having some limitations that can be proceed in further research. This research is applied only on the company online dataset and getting the priority of RFM parameters on the basis of that company behaviors. But this RFM priority may be differ in other industries or trades, company having their own priorities also on the R, F, and M For clustering in this research we used k-mean clustering algorithm, in future other clustering algorithm also used like PAM (partition around mediod) clustering, CLARA (clustering large applications).

REFERENCES

[1] Hosseini, S.M.S., A. Maleki, and M.R. Gholamian, Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. Expert Systems with Applications, 2010. 37(7): p. 5259-5264.

[2] Cheng, C.-H. and Y.-S. Chen, Classifying the segmentation of customer value via RFM model and RS theory. Expert systems with applications, 2009. 36(3): p. 4176-4184.

[3] Kracklauer, A.H., D.Q. Mills, and D. Seifert, Customer management as the origin of collaborative customer relationship management, in Collaborative Customer Relationship Management. 2004, Springer. p. 3-6.

[4] Saarijärvi, H., H. Karjaluoto, and H. Kuusela, Extending customer relationship management: from empowering firms to empowering customers. Journal of Systems and Information Technology, 2013.

[5] Kotler, P., Marketing during periods of shortage. Journal of Marketing, 1974. 38(3): p. 20-29.

[6] Dwyer, F.R., Customer lifetime valuation to support marketing decision making. Journal of Direct Marketing, 1997. 11(4): p. 6-13.

[7] Zhang, S.T., Firm valuation from customer equity: When does it work and when does it fail? International Journal of Research in Marketing, 2016. 33(4): p. 966-970.

[8] Dibb, S. and L. Simkin, Market segmentation: diagnosing and treating the barriers. Industrial Marketing Management, 2001. 30(8): p. 609-625.

[9] Dibb, S., Market segmentation: strategies for success. Marketing Intelligence & Planning, 1998.

[10] Aggarwal, N. and K. Aggarwal, A mid-point based k-mean clustering algorithm for data mining. International Journal on Computer Science and Engineering, 2012. 4(6): p. 1174-1180.

[11] Teknomo, K., K-means clustering tutorial. Medicine, 2006. 100(4): p. 3.

[12] Wu, H.-Y., G.-H. Tzeng, and Y.-H. Chen, A fuzzy MCDM approach for evaluating banking performance based on Balanced Scorecard. Expert systems with applications, 2009. 36(6): p. 10135-10147.

[13] Bozbura, F.T., A. Beskese, and C. Kahraman, Prioritization of human capital measurement indicators using fuzzy AHP. Expert systems with applications, 2007. 32(4): p. 1100-1112.

[14] Pakhira, M.K., S. Bandyopadhyay, and U. Maulik, Validity index for crisp and fuzzy clusters. Pattern recognition, 2004. 37(3): p. 487-501.

[15] Khajvand, M. and M.J. Tarokh, Estimating customer future value of different customer segments based on adapted RFM model in retail banking context. Procedia Computer Science, 2011. 3: p. 1327-1332.

[16] Jackson, S., S.A. Jackson, and D.E. Brashers, Random factors in ANOVA. 1994: Sage.

[17] Goos, P. and D. Meintrup, Statistics with JMP: Hypothesis Tests, ANOVA and Regression. 2016: John Wiley & Sons.

Authorized licensed use limited to: University of New South Wales. Downloaded on October 24,2020 at 12:30:42 UTC from IEEE Xplore. Restrictions apply.