E-ISSN: 2717-2937 | P-ISSN: 2783-1337

International Journal of Research in Industrial Engineering

www.riejournal.com

Int. J. Res. Ind. Eng. Vol. 11, No. 1 (2022) 62–76.

Paper Type: Research Paper

Customer Segmentation to Identify Key Customers Based on RFM Model by Using Data Mining Techniques

Abdolmajid Imani[1], Meysam Abbasi[1], Farahnaz Ahang[1,*], Hassan Ghaffari[1], Mohamad Mehdi[1]

[1] Faculty of Management and Economics, University of Sistan and Baluchestan, Zahedan, Iran; abdolmajidimani@gmail.com; meysam.abbasi13@gmail.com; ahang1989@yahoo.com; hassan_ghaffari@pgs.usb.ac.ir; mohamad.mehdi@ut.ac.ir.

Citation:

Imani, A., Abbasi, M., Ahang, F., Ghaffari, H. & Mehdi, M. (2022). Customer segmentation to identify key customers based on RFM model by using data mining techniques. *International journal of research in industrial engineering, 11* (1), 62-76.

Received: 22/06/2021 Reviewed: 19/07/2021 Revised: 01/10/2021 Accepted: 06/10/2021

Abstract

Accurate identification, attracting, and keeping the customers particularly loyal Customer Relationship Management (CRM) with the goal of optimum allotment of resources and achievement to higher profit is not a competitive profit, but it is a life persistence necessity of companies in virtual space. One of the challenges of companies in this part is how to identify the customer's traits and the separation of different segments of them. Now Customer Lifetime Value (CLV) is the comparison priority in the segmentation of customers to congruous segments. The main goal of this research is to identify key or strategic customers using the RFM model. In this part after determining the amount of Recency, Frequency and Monetary (RFM) in, registered transactions of one store in Iran (Refah Chain Store) at a time about seven months from 23 September 2017 to 20 April 2018 (71161 transactions as final inputs were used), the weight of each variable according to the fuzzy Analytic Hierarchy Process (AHP) was determined. At the next stage customers using the K-means and Two-step's algorithms were clustered and K-means the method according to the Silhouette index was the better algorithm of this letter. According to the results, customers were segmented into three parts and CLV was calculated and for identifying key or strategic customer segmentation, the clustering process was repeated and priorities of all clusters were indicated. Results of data analysis are below: Segment 3: customers of this segment were 3425 members and 11.5% of all company customers were the most loyal customers those are identified as golden customer's segment and all of the variables were higher than average of all data. This research identified the valuable customers for the shop, and it gives them a chance to choose goal customers and invest in them.

Keywords: Customer Lifetime Value (CLV), Customer Relationship Management (CRM), RFM model, Segment.

1 | Introduction

From the time, that strategy of decreasing costs and increasing profits were attention ed to companies, maintaining the customers became one of the strategic goals of the organizations [1]. In this competitive market, satisfying the customer was a complicated issue [2]. It researches showed that a company has a better chance of selling a product to old customers than new ones, as a successful chance of a company for reselling to an active customer is about 60 to 70 percentage, on the other hand, a successful chance for selling to a new customer is about 5 to 20 percentage [3]. Customer Relationship Management (CRM) plays an important role in this regard. CRM focuses on enhancing, maintaining, and establishing long-term associations with customers [4]. CRM is a technology and integrates the process of business that should be coordinated to eliminate the needs of customers and to perform that all the elements within the system should be considered carefully,

Corresponding Author: ahang1989@yahoo.com

http://dx.doi.org/10.22105/riej.2021.291738.1229

and all the expectations should be managed [5]. CRM is attempting to harness new heterogeneous data sources for developing innovative value propositions [6]. Having this data can reply to the organization's knowledge requirement to predict and identify the best behavior with customers and effective service to the different groups of customers can be achieved and totally and the profit van is increased. Data Mining (DM) can help CRM in different ways and easy the process can be achieved [7]. DM was explained as "the process that applies statistical, mathematical, artificial intelligence and machine-learning techniques to extract and identify useful information and subsequently gain knowledge from large databases [8]. In CRM, more information was obtained from DM, which has revealed that not all customers make the same contribution to businesses, and for obtaining a business profit, it is necessary to segment the customers before designing effective marketing strategies [9]. Customer segmentation is done by using behavioral data since it is commonly available and continuously evolving with time and purchase history [10]. Customer Lifetime Value (CLV) is a quantitative measurement that can facilitate customer segmentation. CLV was explained as "the present value of the future profit stream expected over a given time horizon of transacting with the customer [11]. Recency, Frequency, and Monetary (RFM) analysis is one way that can be useful in the CLV concept [12]. RFM is a renowned technique used for evaluating customers based on their buying behavior [13]. RFM analysis nowadays can be conducted using DM methods like two-steps and K-Means [14]. Recent research wants to after weighting 3 variables of RFM model by using of Chang AHP with two-steps and K-Means algorithm, to Customer segmentation of one chain store in Iran Upon ranking the sections based on the CLV, all sectors are calculation and key or strategic customers are identified. Using simple queries in Structured Query Language (SQL) and various common reporting tools, information can be provided to users to conclude the data and the logical relationships between them. When the volume of data is high, users, no matter how professional and experienced, can not distinguish useful patterns among the mass of data, or if they can do so, the cost of operation is very high. Therefore, it can be said that there is currently a paradigm shift from classical modeling and analysis based on basic principles to evolving models and related analyzes directly from the data. DM is one of the most important methods by which useful patterns in data with minimal user intervention are identified. Since the resources of any company are limited, how these resources are allocated to gain a competitive advantage is very important. On the other hand, every company should try to identify more loyal customers in the customer relationship department and allocate more resources to them. In this article, due to the importance of this issue, we tried to show Refah Chain Store how they can use the DM technique and increase their competitive advantage. A review of the research background shows that no research in this collection (Refah Chain Store) has used DM techniques to segment customers and this is an innovation in the present study.

2 | Literature

2.1 | Customer Relationship Management (CRM)

In today's global competition, a variety of products must be made available to the customer according to the request [15]. It should be said that successful businesses try harder to satisfy customers [16]. Relationship Marketing (RM) principle is a suitable zone for modern-day marketing, and CRM is the root of that [17]. Although the appearance of CRM as an important issue at a job and career is known there is no accepted definition for this [18]. Nowadays, CRM leads to a holistic issue for improving the management of long-term profitable customer relationships, and with this method, following the business processes and customer loyalty can be achieved [19]. For creating a fantastic CRM in an organization, all the equipment and issues should be controlled. CRM as a piece of equipment with high technology of web/app makes an organization capable of comprehending customers or potential customers, and this matter helps the customers to make a suitable decision and creates good transactions [20]. Some definition of CRM is offered. CRM is a commercial strategy and marketing that synthesis the technology process and all activities of a career [21]. Essay emphasis that new definitions of the importance of CRM as a strategic process for the maximum value of a customer for the organization. Ballantyne [22] the concept of CRM has related to the development of an organization that requirements and preferences of the customer have another definition. Gummesson [23] defined that the basis of CRM is RM logic and codification of goals

Imani et al.| Int. J. Res. Ind. Eng.11(1) (2022) 62-76

and strategies of that. CRM is meaning work to manage relationship management on a big scale with long profit persistence in the customer's mind. CRM is defined as an intensive career strategy on customers and its goal is to increase the customer's satisfaction and loyalty [24]. Nowadays, marketing managers distinguish is concentrated on CRM on stale and long relationships with the customer, and it is valuable for both sides [25]. CRM is divided into three parts [26]:

Operational CRM. In this method, all the stages of the relationship with the customer, from marketing to selling, after-sales services, and the feedback from customers are surrendered to one person. However, in a way, that salesmen and engineers have the history of the customers without their presence.

Analytical CRM. At this part, some instruments and issues are used to analyze the information from analytical CRM and prepare the results for commercial operation management. Analytical CRM and operating CRM are in a bilateral interaction, which means that operating data are analyzed. After analysis, results have a direct effect on the operating part. It means that by helping the analysis of this part, the customers were classified and the organization concentrates on a special part.

Collaborative CRM. At this part relationship with the organization is possible by phone, mobile, fax, the Internet or, etc. Interaction CRM is more efficient because of choosing the issue by customers and minimum time of the process (from collecting the data to process and delivered to the customer).

2.2 | DM Techniques

DM is the art of finding information or knowledge in a large amount of data. Like statistics, DM is becoming increasingly common in companies and organizations that want to extract relevant information from their databases, which they can use for their own needs [27]. The DM technique has no standard terminology, but the most popular terms are algorithm and technique. While the word technique describes a wide range of procedures employed while conducting DM, algorithms go into more detail. However, they are both used for the same concept. Aghimien et al. [28] opined that the DM techniques include; market-based analysis, memory-based reasoning, cluster detection, link analysis, decision trees, and rule induction, artificial neural networks, genetic algorithms, and online analytic processing. All DM techniques are divided into two large groups according to the principle of working with the original data. The upper level of this classification is determined based on whether the data DM after they are stored or distilled for later use:

– *The direct use of the data, or save the data: In this case, the original data is stored in a detailed and explicit directly used for predictive modeling steps and/or analysis exceptions. The problem with this group of methods is that their use can be difficult analyzing very large databases. Method's nearest neighbor method and k-nearest neighbor belong to this group.*
– *Identify and use formal laws or distillation templates: When distillation technology templates one sample (template) information is extracted from the raw data is converted into a certain formal structure whose form depends on the DM method. This process is performed to the stage of the free search. Note that the first group of methods, basically, this step absent [29].*

2.3 | K-Means Algorithm

The goal of the K-means algorithm is to divide M-points in N-dimensions into K-clusters to minimize the within-cluster sum of squares "Local" optima solution was sought so no movement to a point from one cluster to another will reduce the within-cluster sum of squares the algorithm requires as input a matrix of M-points in N-dimensions and a matrix of K-initial cluster centers in N-dimensions. The number of points in cluster L is denoted by NC(L). D (I, L) is the Euclidean distance between point I and cluster L. The general procedure is to search for a K-partition with the locally optimal within-cluster sum of squares by moving points from one cluster to another [30].

2.4 | Two-Step Analysis

Two-step analysis requires only one pass of data (which is important for very large data files), and it can produce solutions based on mixtures of continuous and categorical variables. The algorithm is based upon a distance measure that gives the best results if all variables are independent, continuous variables have a normal distribution, and categorical variables have a multinomial distribution. This is seldom the case in practice, but the algorithm is thought to behave reasonably well when the assumptions are not met [31].

2.5 | Customer Segmentation

Dividing the customers is a way of classifying customers into different branches [32]. For improving the financial matters and business customer segmentation is a suitable and important way. The clustering algorithm can help authorities to make this target possible [33]. This action is a balance between non-identification of customers and one by one identification of customers who can be caused of targeting the activities of CRM, marketing management, and allocating marketing resources, in another hand, it is advantageous for identifying the customers one by one especially for the time that one organization work with many customers. Customer segmentation was done according to customer's requests in the past, whereas in recent years, by changing the organization procedure from concentrating on the product as a value creator factor to concentrating on the customer as a value producer fund, customers are segmented according to their value amount [34]. Subtly segmentation causes organizations can know profitable customers, understand their customer's requests, allocate resources, and be against rivals [35]. According to Lemon and Mark, there is no special idea for customer segmentation. The best segmentation model is that, draws an appropriate intuition from customers for managers first, and second, helps to effective achievement to market and finding out suitable answers from customers [36].

2.6 | Customer Lifetime Value (CLV)

Different definitions for value involve concepts, utility, benefits, and emotional bonds that indicated different segments of value [37]. CLV is a core metric in CRM. It can be effective for improving market segmentation, allocating resources, analyzing the rival company, customize marketing communication, improving the timing of product exposure, and nominate the value of a company [38]. Dynamism is a trait for a valuable customer and after a while relationship with him is changed, so comprehending the relationship that is creating by investigating the life circle of customers is the major inscription of CRM [39]. Move for customer satisfaction at marketing and also increasing access to data of customers transaction creates an interest in customers about understanding and analysis of CLV concept [40]. CLV is analyzed with titles like customer value, CLV, customer's right, and customers profit at studies [41]. Existing variable definitions of CLV. It shows different issues and attitudes about this matter. Some definitions about CLV are below: CLV is defined as a profit that a customer creates for a company at the time and emphasizes a long relationship with a customer and defines that a company should not pay for a customer more than the profit is prepared from him [40]. Researchers define this concept like this, set of the values that customer creates for an organization at the time and calculating of that can be caused by profit and optimum allotment of resources [42]. CLV depends on all transactions that a customer is creating in the future [43]. CLV is the sum of pecuniary circulations that are created at A CLV or part of customers [44]. The most common issues for the determination of CLV are Net Present Value (NPV), Share of Wallet (SOW), Markov chain, Return on Investment (ROI), and RFM. RFM model is the most common [42], Therefore, in this research, we use variables RFM models to analyze customers and calculate the CLV.

2.7 | RFM Model

One of the simplest and the most powerful model for CRM is the RFM model [45]. RFM model is a famous and applied model for analyzing the customer value, and the best point for that is extracting the customer's traits with lower priority and by helping the segmentation method [46]. Some of the quantitative traits are discovered by the RFM model and the variables for 15 potential relationships in CRM are enriched by this model too because the customer value can be considered by novelty utilization, time of shopping in casual utilization, and monetary cost of consumers in the model [47]. The variables of this model are defined below [46].

The recency of the last purchase (R). This variable is pointing at recency, the interval between the last purchase by the customer to the end of an especial period. The shorter interval shows the higher value of this variable in the model.

Frequency of the purchases (F). This variable shows exchanging several transactions. It is the number of transactions that a customer is going through at a special period. A higher number shows a higher variance in this model.

The monetary value of the purchases (M). This variable shows purchase monetary value. It is the money amount that a customer spent at a special period for purchase. Higher spent money amount shows the higher variable in this model.

Hughes [48] defines that these three variables have identical importance, so the weight for each three of them is specific and identical. In another study, Stone [49] believes that because of different traits in every industry, three variables will have different importance. He determines the weights of RFM index assurance in his study.

Imani et al.| Int. J. Res. Ind. Eng.11(1) (2022) 62-76

3 | Fuzzy Analytic Hierarchy Process (AHP)

Fuzzy logic has gained wide acceptance in the field of accounting and business. This acceptance is due to the ability to manage in situations of ambiguity and lack of consistency that does not exist within other approaches to dual value logic [50]. Fuzzy logic tools help us to have a more accurate estimate in the future [51]. Analytic Hierarchy Process (AHP) is a kind of multi-index instrument that is used for deciding and provides ways to solve complicated problems in a hierarchy behavior, and also it is used for analyzing the indexes which are related to suggestions that must be given [52]. All the superseded are compared two by two according to suitable priority and a list of superseded are collected for each priority scale [53]. Preference scales are effective, and the most used about them are the scales between "equal importance" and "extreme importance." By using fuzzy AHP decision experts are a cable for giving pragmatic scores for superseded in a case that lots of doubts exist. About various outlook's different kinds of fuzzy AHP can be participated. One of the important models for analyzing is Chang's model and the thing which is important for this model is the degree of the possibility of each index. For developing the pairwise comparison scale, numbers of fuzzy triangular play important roles, and at each level of the hierarchy pair, a wise comparison matrix is created. Then a proper subset of each row to the matrix is calculated respectively to have a new collection in a hierarchy. Overall triangular fuzzy values (li, mi, and ui) for Mi index are gotten by calculating Ii/*Ii, mi/*mi, and ui/*ui, (i=1, 2, …, n). By using these values fellowship roles are calculated for each index. Normalization was done and the final important weights of each index are obtained. According to Chang's issue, to apply the process that is dependent on this hierarchy, extent analysis for each index is performed and extent analysis for each index gin is done respectively. So, for calculating the M content values about each index following *Eq. (1)* was used:

$$S_i = \sum_{i=1}^{m} M_{g_i}^{j} \otimes \left[\sum_{i=1}^{n} \sum_{j=1}^{m} M_{g_i}^{j}\right]^{-1}. \quad (1)$$

In a situation that gi is the target collection (i= 1, 2, 3, …, n) and all the gi M (j= 1, 2, 3, 4, 5, … m) are Triangular Fuzzy Numbers (TFNs). Different steps of Chang's model are shown as in bellow:

- *Eq. (1) expresses the fuzzy synthetic extent value (Si) concerning the Ith index.*
- *Eq. (2) defines the degree of possibility of M2= (I2, m2, u2) Mi = (I1, M1, u1).*

$$V(M_2 \geq M_1) = \sup_{y \geq x}\left[\min\left(\mu_{M_1}(x), \mu_{M_2}(y)\right)\right]. \tag{2}$$

- *x and y are values on the diagram of each index fellowship role. This formula can be defined as obtained in Eq. (3).*

$$V(M_2 \geq M_1) = \begin{cases} 1, & \text{if } m_2 \geq m_1 \\ 0, & \text{if } I_1 \geq u_2 \\ \frac{I_1 - u_2}{(m_2 - u_2) - (m_1 - I_1)} & \text{otherwase} \end{cases}. \tag{3}$$

4 | Experimental History

Tleis et al. [54] clustering the organic food market in their study in Lebanon. For doing this research 320 questionnaires between consumers of organic food at Beirut were distributed. By analyzing the questionnaire, consumers by using of k- means algorithm was clustered to four branches with name's health-conscious, localist, irregular and rational and for each cluster of them, suitable strategies were identified. Hiziroglu and Senbas [55], in their research, use fuzzy clustering in the automotive industry of customers. The goal of this research was clustering using the fuzzy C-means algorithm and comparing them with the algorithm of traditional clustering. For this research, the data of 130 customers of the automotive supplier was received. Results showed more balanced clustering than the traditional one and help marketing managers to understand their customers more. Alhilman et al. [56] because of increasing the profit and customer loyalty and preventing them not to take them apart segment them at Indonesia and for identifying PDX of communication company customers, service and suitable service for every segment some suggestions were offered. Data for payment and liability was analyzed. According to this segmentation, the SPSS package of customers with different suggestions like increasing the customer's information, services, mutual sale, and more commercial was offered. Wang and Lu [57] do research using DM for relationship management modeling with China's communication company customers and segment them according to their requests. Customer's data at model were such as habits, requirements, sex, and income amount related to K-means the center. At this research, algorithms were used and users were divided into three segments: key, important, and public. Dorsan and Caber [58] using RFM analysis segment the customers of three five-star hotels In Antalya turkey. Analysis shows that 369 profitable were divided into eight groups: loyal summer season customers, collective buying customers, winter season customers, lost customers, high-potential customers, new customers, winter season-high, potential customers, new customers, and winter high potential customers and many of them were at gone customer's segment (36%). Results of research help the hotel's manager to adopt strategies of relationship with the customer. Wei et al. [59] by combining DM K-Means, SOM and using the RFM model for segmenting the customers in the hairdresser in Taiwan. Results showed four kinds of customers: loyal customers, potential customers, new customers, and lost customers, and for every kind of them, marketing strategies were identified. Miguéis et al. [60] segment the customers of food retail in Europe, they analyze shopping and product segmentation by using a hierarchy algorithm and customer's attribution to a special branch according to their shopping history. Research results showed an improvement in the relationships of the company and customers by marketing. Liao et al. [61] by using K-Means and Apriori algorithm segmented the customers of the Carrefour hypermarket in Taiwan. With using branch analysis, they divided the customers and a list of customer's requests was designed. With using this method, for more attractions of the list, online shopping and delivery at home were offered for customers. Imani and Abbasi [62] analyzed customers of a Chain store in Iran using the RFM model. To carry out this research, they used the process of the fuzzy AHP and fuzzy C-Means algorithm, and customers were divided into seven clusters that provided a framework for developing customer relationship strategies.

5 | Research Framework

The main goal of this research is segmenting and analyzing the customer's behavior according to behavior transaction data, to be determined the key or strategic customers. The total framework of this research is drawn in *Fig. 1* and shows phases cluster and store customer ranking according to their CLV. At these research variables of the RFM model, fuzzy AHP, K-means, and Two-step algorithm were used.

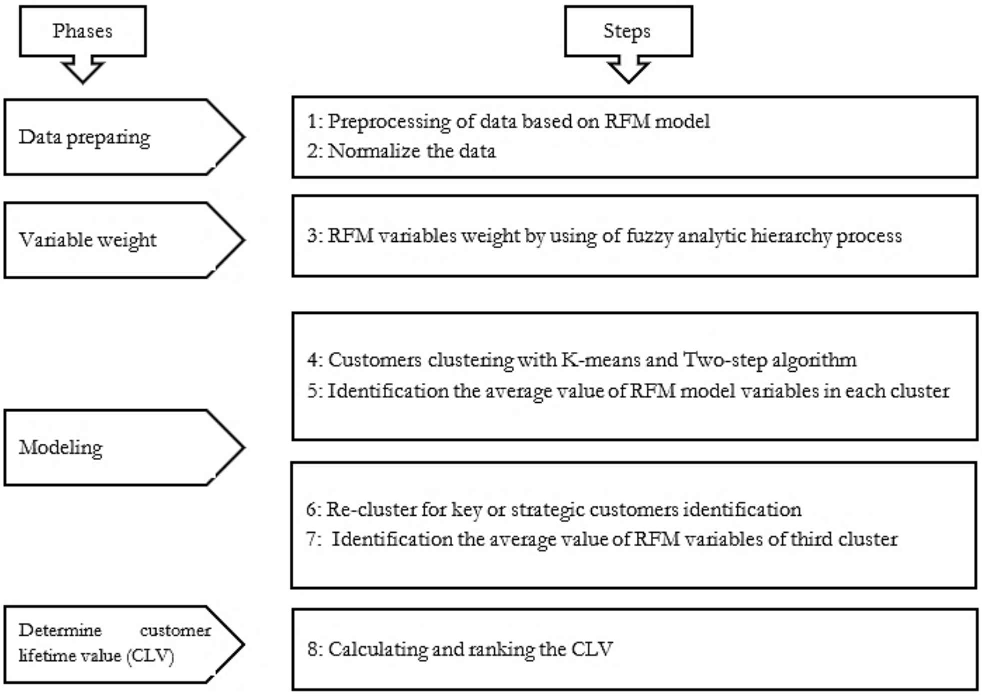

Fig. 1. Research framework.

The computing process can be discussed in detail in the next section.

6 | Finding

6.1 | Phase 1: Data Preparation

Step 1. Preprocessing of data based on the RFM model: in recent research, registered transactions of one store in Iran at a period of time of about seven months from 23 September 2017 to 20 April 2018 was used. After receiving the data and doing the preparation process, 71161 transactions as final inputs were used. The preparation process has two steps. At first, step purging the data was done, and some data with invalid amounts were identified and omitted. At the second step, RFM model variables using SPSS Modeler 18 software were calculated. *Table 1* shows part of the prepared data for clustering.

Table 1. Calculating the variables amount of RFM.

Customer Number	Recency (R)	Frequency (F)	Monetary (M)
001	18	3	1900317
002	55	6	2897889
003	98	15	2849974
.	.	.	.
.	.	.	.
.	.	.	.
29785	3	2	335100

Step 2. Normalize the data: because of differences at the RFM model variables unit. These amounts should be normalized as an identical unit, so these amounts were normalized by the Min-Max normalization method.

$$NR = \frac{R_{max} - R}{R_{max} - R_{min}}, \qquad NF = \frac{F - F_{min}}{F_{max} - F_{min}}, \qquad NM = \frac{M - R_{min}}{M_{max} - M_{min}}. \tag{4}$$

Eq. (4) shows the Rmax, Fmax, and Mmax the highest value and Rmin, Fmin and Mmin shows the lowest value of variables, and NM, NF and, NR show a normal number of variables. If M and F variables where a higher customer situation is more optimal and if the R variable is the lower customer has a better situation. *Table 2* shows the normalized amounts of data.

Table 2. Normalized value of data.

Cluster Number	NR	NF	NM
001	0.9143	0.0364	0.0943
002	0.7381	0.0909	0.1441
003	0.5333	0.2545	0.1417
.	.	.	.
.	.	.	.
.	.	.	.
29875	0.9857	0.0182	0.0161

6.2 | Phase 2: Variable Weight

Step 3. RFM variable's weight by using the fuzzy AHP: at the next step identification of the RFM model variable's weight with using of the fuzzy hierarchical analytical process is set. So asked the store managers to Pairwise comparison judgments between RFM variables. *Table 3* shows the arithmetic average of the expert people's ideas. In the last column of this table, some of the elements of the row are shown.

Table 3. Matrix average of expert people ideas.

Variable	R	F	M	SUM	Normalized
R	(1,1,1)	(1,0,583,0,583)	(0,5,0,5,1)	(2,.83,2,.83,3)	(0,189,0,21,0,379)
F	(1,1,75,1,75)	(1,1,1)	(0,833,0,833,1)	(2,833,3,583,3,75)	(0,258,0,316,0,474)
M	(1,2,2)	(1,1,25,1,25)	(1,1,1)	(3,4,25,4,25)	(0,273,0,429,0,537)
SUM				(7,916,9,916,11)	
CRm =0.021		CRg =0.028			
Compatible					

For the determination of compatibility, Gogus and Boucher's method was used because the numeral average matrix and middle number matrix were calculated lower than 0.1 and showed the people's preferences. And in continuation, the final higher degree normalized weights of each. Variables are calculating, and their results are shown in *Table 4* and *Fig. 2*.

Table 4. Calculating preference degree and normalized weight.

Variables	R	F	M	Preference Degree	Normalized Weight
R	-	0.445	0.327	0.327	0.158
F	1	-	0.749	0.749	0.316
M	1	1	-	1	0.482
SUM				2.076	1

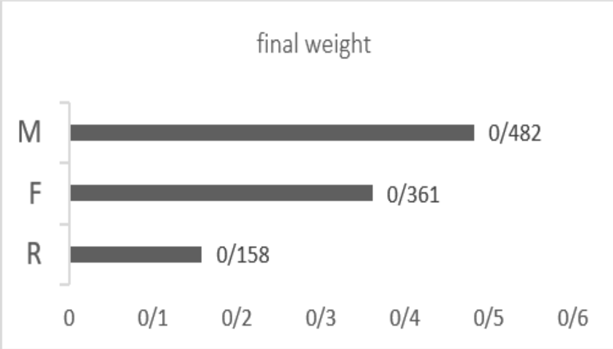

Fig. 2. Final weight of each RFM model variables.

Table 5. Data weighted values.

Cluster Number	NR*WR	NF*WF	NM*WM
001	0.1445	0.0131	0.0454
002	0.1166	0.0328	0.0695
003	0.0843	0.0919	0.0683
.	.	.	.
.	.	.	.
.	.	.	.
29875	0.1557	0.0066	0.0078

6.3 | Phase 3: Modeling

Step 4. Customer's clustering with K-Means and Two-step Algorithm step: at recent research K-Means and a Two-step algorithm for clustering around the customers were used. So, SPSS Modeler 18 that is one of the most powerful software for DM was used. Analyzing the Clustering quality and superiority rate of one cluster in comparing with other clusters with different clustering algorithms or identical algorithms but with different parameters are different and at this research Silhouette index. Silhouette separation and density index are shown with good, average, and weak amounts. Average of the Silhouette index amount for evaluation of clustering validity and also deciding about choosing optimal clusters is used and that the rate is calculating according to distance and proximity of observations and Clusters to each other, and it is calculating by this *Eq. (5)*:

$$S(i) = \frac{(b(i)-a(i))}{\max(a(i).b(i))}. \quad (5)$$

In *Eq. (5)* an (i) is the distance average between observations, i with other observations at one identical cluster, and b(i) is the observation distance averages, i to all observations at other clusters. According to the top formula amount of S(i) is between -1 and +1. If S(i) is nearer to +1, it means that sample clustering has been done as well as possible, and offered to cluster for a sample is suitable, but if S(i) is nearer to -1, it means that sample clustering has not been done well, and suggested clustering for data is not suitable According to store manager's ideas, number of clusters were clustered between 3 to 10.

As it is indicated in the table, the Silhouette amount of the K-means algorithm has the highest amount, and it shows that the Clustering quality is better, so for the Clustering of store customers these algorithms are used.

Table 6. Silhouette index values for K-means and two-step algorithm.

Cluster	Silhouette Index for Two-Step	Silhouette Index for K-Means
3	0.123	0.613
4	0.084	0.579
5	-0.031	0.565
6	-0.072	0.555
7	0.053	0.535
8	0.034	0.543
9	0.016	0.517
10	-0.003	0.518

Step 5. Identification of the average value of RFM model variables for each cluster: the average of each variable at every cluster with the dividing of variable value at that cluster by customer's number based on the below relationship is calculated.

$$\bar{R} = \frac{\sum R}{n}, \qquad \bar{F} = \frac{\sum F}{n}, \qquad \bar{M} = \frac{\sum M}{n}. \tag{6}$$

Table 7. Details clustering according to K-means algorithm.

Cluster	Average R	Average F	Average M	Percent of Customer	Number of Customers
1	0.0392	0.0018	0.0255	29.7%	8857
2	0.1199	0.0067	0358.0	58.8%	17503
3	0.1406	0.0407	0.1923	11.5%	3425
Average	0.0983	0.0091	0.0507		

For the determination of key customers calculating the CLV is done. By summing the value average of 3 RFM, CLV is brought, and ranking the clusters is done.

Table 8. Calculating of CLV.

Cluster	CLV	CLV Rank
1	0.0665	3
2	0.1624	2
3	0.3736	1

As indicated in the table, eight third segments with 0.3736 CLV amount have the most loyal customers.

Step 6. Re-cluster for key or strategic customers identification: According to *Table 8* and the main goal of this research is to identify key customers. Third customer clusters are again clustered with K-means and Two-step algorithm. *Table 9* shows the results and details of clustering. As it is indicated, when the customers are divided into three clusters, the Silhouette index has the highest amount, and it is chosen as the optimal amount for the replacement cluster.

Table 9. Silhouette index amount for K-Means and two-step algorithm for the third cluster.

Cluster	Silhouette Index Two-step	Silhouette Index K-Means
2	0.408	0.496
3	0.408	0.543
4	0.386	0.53
5	0.424	0.317
6	0.393	0.335
7	0.377	0.359
8	0.323	0.345
9	0.299	0.353
10	0.301	0.324

Step 7. Identification of the average value of RFM variables of the third cluster: Using *Eq. (6)*, the average values on each variable are calculated for new clusters.

Table 10. Average value of RFM variable of the third cluster.

Cluster	Average R	Average F	Average M	Percent of Customer	Number of Customers
3-1	0.12	0.0302	0.2025	35.5%	1217
3-2	0.1521	0.0726	3139.0	16.8%	575
3-3	0.1520	0.0372	0.1418	47.7%	1633
Average	0.1406	0.0407	0.1923		

6.4 | Phase 4: Determine CLV

Step 8. Calculating and ranking the CLV.

Table 11. Calculation of CLV cluster ranking.

Cluster	CLV	CLV Rank
1	0.0665	5
2	0.1624	4
3-1	0.3527	2
3-2	0.5386	1
3-3	0.331	3

Companies can use their limited resources for beneficial and valuable customers by calculating the CLV for every cluster, as you observe from a table, five customers of clusters 1 to 3 are perched at first rank, and this cluster has the best customers.

7 | Conclusion

DM can forecast the potential customer profitability that changes to actual and shows the period of customer loyalty and how they will leave the company. With increasing the importance of CRM in today's commercial environment, a lot of organizations concentrate on topics that are related to customer identification, loyalty, and profitability. Recent research was done with the goal of key or strategic customer identification based on variables of the RFM model and Two-steps and K- means algorithm. Results of data analysis are below: Segment 3: customers of this segment were 3425 members and 11.5% of all company customers were the most loyal customers those are identified as golden customer's segment and all of the variables were higher than average of all data. This RFM is related to the segment model formed from three subsidiary segments that are titled: golden customer's segment of degree 1, golden customer's segment of degree 2, and degree 3. Between the 3 segments, the golden customer's segment of degree 1 had CLV 0.5386, and 16.8% of third segment customers were identified as valuable customers of the company. CLV of golden customer's segment of degree 2 and 3 respectively was 0.3527 and 0.331 and formed respectively 35.5% and 47.7% of third clusters customers, and they are taken place at second and third rank. Company services to these customers shouldn't be limited to their casual programs, but some special programs should be considered for these three segments so the company should pay more costs for these customers. Segment 2: customers of this segment are known as silver customers and the CLV of this segment is 0.1623, and they are formed 58.8% of all customers. The recency of this segment is higher but the frequency monetary is lower than the total average so a suitable politic should be performed for encouraging the customers to be joined to the golden customers. Segment 1: the customers of segment 1 are bronze company customers, and they were 29.7% of all companies' customers, and their CLV was 0.666 and all of the RFM variables were lower than the total average. For performing the politics of relationship with customers from customers of this segment from the company a suitable aligns should be considered between relationship costs with the customer with the income that is created by customers. This research identified the valuable customers for the shop, and it gives them a chance to choose goal customers and invest in them, but it doesn't mean that the shop doesn't pay attention to the other customers, and they limited their effort to satisfy their profitable customers but it means that they allocated a suitable budget for performing the relationship programs with customers to increase their loyalty and satisfaction. Using DM in CRM can be very useful in the following cases:

- *Helps in sales forecast: DM can help predict future trends by analyzing people's past behaviors.*
- *Helps in market segmentation: DM helps to properly segment your target audience based on demographic information, shopping behavior, gender, and other factors. Information can be gathered through market research, social media platforms, and more.*
- *Helps to make quick and smart business decisions: DM uses predictive model analysis to determine the lifetime value of each customer. With such information and in-depth insight, it allows creating personalized services for each customer with the assurance of proper budget allocation.*
- *Helps to detect fraud: For starters, it analyzes past fraudulent activities to prevent repetition.*
- *Helps to increase customer loyalty: For example, DM uses a model called "customer cluster", which uses audience data on social media sites to generate ideas for improving brand service, customer satisfaction, and increasing loyalty.*

Given all of the above, the use of DM techniques is essential for any business in today's world. Because businesses need to improve their competitive advantage to succeed. And using DM techniques in CRM can help to gain a competitive advantage. One of the businesses of our society is Refah Chain Store that communicates with many customers every day, which can use the data of these customers and achieve a competitive advantage by formulating appropriate strategies. Also, the use of this technique in chain stores leads to the attention and importance of this issue in such businesses, because so far, no research has examined DM in chain stores.

Reference

[1] Kuo, Y. F., Wu, C. M., & Deng, W. J. (2009). The relationships among service quality, perceived value, customer satisfaction, and post-purchase intention in mobile value-added services. *Computers in human behavior*, *25*(4), 887-896.

[2] Hu, Y. H., Huang, T. C. K., & Kao, Y. H. (2013). Knowledge discovery of weighted RFM sequential patterns from customer sequence databases. *Journal of systems and software*, *86*(3), 779-788.

[3] Griffin, J., & Lowenstein, M. W. (2002). *Customer winback: how to recapture lost customers--and keep them loyal*. John Wiley & Sons.

[4] Soltani, Z., Zareie, B., Milani, F. S., & Navimipour, N. J. (2018). The impact of the customer relationship management on the organization performance. *The journal of high technology management research*, *29*(2), 237-246.

[5] Ghalenooie, M. B., & Sarvestani, H. K. (2016). Evaluating human factors in customer relationship management case study: Private banks of shiraz city. *Procedia economics and finance*, *36*, 363-373.

[6] Mozaffari, A., Bonyadi, A., & Alizadeh, F. (2010). Clustering and discovering the damage patterns of the third-party insurance customers by using data mining techniques. *The 10th international conference on industrial engineering*.Tehran. https://civilica.com/doc/284121/

[7] AbuAli, A. N., & Abu-Addose, H. Y. (2010). Data warehouse critical success factors. *European journal of scientific research*, *42*(2), 326-335.

[8] Ngai, E. W., Xiu, L., & Chau, D. C. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert systems with applications*, *36*(2), 2592-2602.

[9] Hu, Y. H., Huang, T. C. K., & Kao, Y. H. (2013). Knowledge discovery of weighted RFM sequential patterns from customer sequence databases. *Journal of systems and software*, *86*(3), 779-788.

[10] Harwati, A. M., & Karunia, A. (2017). Utilization of social media for consumer behavior clustering using text mining method. *Journal of engineering and applied sciences*, *12*(3), 6406–6411. http://docsdrive.com/pdfs/medwelljournals/jeasci/2017/6406-6411.pdf

[11] Safari, F., Safari, N., & Montazer, G. A. (2016). Customer lifetime value determination based on RFM model. *Marketing intelligence & planning*, 34(4), 446–461.

[12] Jafari Momtaz, N., Alizadeh, S., & Sharif Vaghefi, M. (2013). A new model for assessment fast food customer behavior case study: an Iranian fast-food restaurant. *British food journal*, *115*(4), 601–613.

[13] Christy, A. J., Umamakeswari, A., Priyatharsini, L., & Neyaa, A. (In Press). RFM ranking–An effective approach to customer segmentation. *Journal of king saud university-computer and information sciences*. https://doi.org/10.1016/j.jksuci.2018.09.004

[14] Madeira, S., & Sousa, J. M. (2002, July). Comparison of target selection methods in direct marketing. In *European symposium on intelligent technologies, hybrid systems and their implementation on smart adaptive systems*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.197.2187&rep=rep1&type=pdf

[15] Shababi, H., Yahyazadefar, M., Ghiasabadi Farahani, M., Motamedi, R., & Ghafari Ashtiani, P. (2021). Identification and prioritization of critical success factors in telecommunication supply chain management in Mazandaran province using DANP. *Innovation management and operational strategies*, *2*(1), 30-54.

[16] Sadat Hedayati, T. (2020). The relationship between the factors affecting customer satisfaction for dairy products with path analysis (case study: Golestan pegah company). *Innovation management and operational strategies*, *2*(1). (**In Persian**). https://www.sid.ir/en/journal/ViewPaper.aspx?id=790893

[17] Rahimi, R., & Kozak, M. (2017). Impact of customer relationship management on customer satisfaction: the case of a budget hotel chain. *Journal of travel & tourism marketing*, *34*(1), 40-51.

[18] Ngai, E. W. T. (2005). Customer relationship management research (1992-2002) an academic literature review and classification. *Marketing intelligence & planning*, *23*(6), 582–605.

[19] Pedron, C. D., Picoto, W. N., Dhillon, G., & Caldeira, M. (2016). Value-focused objectives for CRM system adoption. *Industrial management & data systems*, *116*(3), 526–545.

[20] Anshari, M., Almunawar, M. N., Lim, S. A., & Al-Mudimigh, A. (2019). Customer relationship management and big data enabled: personalization & customization of services. *Applied computing and informatics*, *15*(2), 94-101.

[21] Anton, J., & Hoeck, M. (2002). *E-Business customer service*. Santa Monica.

[22] Ballantyne, D. (2003). A relationship-mediated theory of internal marketing. *European journal of marketing*, 37(9), 1242–1260.

[23] Gummesson, E. (2004). Return on relationships (ROR): the value of relationship marketing and CRM in business-to-business contexts. *Journal of business & industrial marketing*, *19*(2), 136-148.

[24] Seeman, E. D., & O'Hara, M. (2006). Customer relationship management in higher education: Using information systems to improve the student-school relationship. *Campus-wide information systems*, *23*(1), 24–34.

[25] Nguyen, T. H., Sherif, J. S., & Newby, M. (2007). Strategies for successful CRM implementation. *Information management & computer security*, *15*(2), 102–115.

[26] Ye, N. (Ed.). (2003). *The handbook of data mining*. CRC Press.

[27] Rémy, N. M., Martial, T. T., & Clémentin, T. D. (2018). The prediction of good physicians for prospective diagnosis using data mining. *Informatics in medicine unlocked*, *12*, 120-127.

[28] Aghimien, D., Aigbavboa, C., & Oke, A. (2019). A review of the application of data mining for sustainable construction in Nigeria. *Energy procedia*, *158*, 3240-3245.

[29] Vadim, K. (2018). Overview of different approaches to solving problems of data mining. *Procedia computer science*, *123*, 234-239.

[30] Telgarsky, M., & Vattani, A. (2010). Hartigan's method: k-means clustering without voronoi. *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 820-827). JMLR Workshop and Conference Proceedings.

[31] Tekień, A., Gutkowska, K., Żakowska-Biemans, S., Jóźwik, A., & Krotki, M. (2018). Using cluster analysis and choice-based conjoint in research on consumers preferences towards animal origin food products, theoretical review, results and recommendations. *Animal science papers & reports*, *36*(2), 171-184.

[32] Khalili-Damghani, K., Abdi, F., & Abolmakarem, S. (2018). Hybrid soft computing approach based on clustering, rule mining, and decision tree analysis for customer segmentation problem: real case of customer-centric industries. *Applied soft computing*, 73, 816-828.

[33] Qadadeh, W., & Abdallah, S. (2018). Customer's segmentation in the insurance company (TIC) dataset. *Procedia computer science*, *144*, 277-290.

[34] Liu, D. R., & Shih, Y. Y. (2005). Integrating AHP and data mining for product recommendation based on customer lifetime value. *Information & management*, *42*(3), 387-400.

[35] Blocker, C. P., & Flint, D. J. (2007). Customer segments as moving targets: integrating customer value dynamism into segment instability logic. *Industrial marketing management*, *36*(6), 810-822.

[36] Lemon, K. N., & Mark, T. (2006). Customer lifetime value as the basis of customer segmentation: Issues and challenges. *Journal of relationship marketing*, *5*(2-3), 55-69.

[37] Zhang, T. C., Gu, H., & Jahromi, M. F. (2019). What makes the sharing economy successful? an empirical examination of competitive customer value propositions. *Computers in human behavior*, *95*, 275-283.

[38] Dahana, W. D., Miwa, Y., & Morisada, M. (2019). Linking lifestyle to customer lifetime value: an exploratory study in an online fashion retail market. *Journal of business research*, *99*, 319-331.

[39] Olafsson, S., Li, X., & Wu, S. (2008). Operations research and data mining. *European journal of operational research*, *187*(3), 1429-1448.

[40] Benoit, D. F., & Van den Poel, D. (2009). Benefits of quantile regression for the analysis of customer lifetime value in a contractual setting: an application in financial services. *Expert systems with applications*, *36*(7), 10475-10484.

[41] Hwang, H., Jung, T., & 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*, *26*(2), 181-188.

[42] Buttle, F., & Maklan, S. (2009). *Customer relationship management: concepts and technology*. Routledge

[43] Glady, N., Baesens, B., & Croux, C. (2009). Modeling churn using customer lifetime value. *European journal of operational research*, *197*(1), 402-411.

[44] Villanueva, J., & Hanssens, D. M. (2007). *Customer equity: measurement, management and research opportunities*. Now Publishers Inc.

[45] Carrasco, R. A., Blasco, M. F., & Herrera-Viedma, E. (2015). A 2-tuple fuzzy linguistic RFM model and its implementation. *Procedia computer science*, *55*, 1340-1347.

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

[47] Song, M., Zhao, X., E, H., & Ou, Z. (2016). Statistic-based CRM approach via time series segmenting RFM on large scale data. *Proceedings of the 9th international conference on utility and cloud computing* (pp. 282-291). https://doi.org/10.1145/2996890.3007873

[48] Hughes, A. M. (1994). *Strategic database marketing: the masterplan for starting and managing a profitable* (2nd Eds.). McGraw-Hill.

[49] Stone, B. (1995). *Successful direct marketing methods NTC business books* (8th Eds.). Lincolnwood, IL.

[50] Imeni, M. (2020). Fuzzy logic in accounting and auditing. *Journal of fuzzy extension and applications*, *1*(1), 69-75.

[51] Kaviani, M., & Fakhrehosseini, S. F. (2020). Application of fuzzy logic in portfolio management: evidence from Iranian researches. *Journal of fuzzy extension and applications*, *1*(2), 115-118.

[52] Saaty, T. L. (1980). *The analytic hierarchy processes*. Mcgraw Hill, New York.

[53] Taha, H. A. (2013). *Operations research: an introduction*. Pearson Education India.

[54] Tleis, M., Callieris, R., & Roma, R. (2017). Segmenting the organic food market in Lebanon: an application of k-means cluster analysis. *British food journal*, 119(7), 1423–1441.

[55] Hiziroglu, A., & Senbas, U. D. (2016). An application of fuzzy clustering to customer portfolio analysis in automotive industry. *International journal of fuzzy system applications (IJFSA)*, *5*(2), 13-25.

[56] Alhilman, J., Rian, M. M., Marina, W., & Margono, K. (2014). Predicting and clustering customer to improve customer loyalty and company profit. *2nd International conference on information and communication technology (ICoICT)* (pp. 331-334). IEEE.

[57] Wang, J., & Lu, H. M. (2010). Application of data mining technology in CRM. *Journal of beijing information science & technology university*, *25*. https://en.cnki.com.cn/Article_en/CJFDTotal-BJGY201002019.htm

[58] Dursun, A., & Caber, M. (2016). Using data mining techniques for profiling profitable hotel customers: an application of RFM analysis. *Tourism management perspectives*, *18*, 153-160.

[59] Wei, J. T., Lee, M. C., Chen, H. K., & Wu, H. H. (2013). Customer relationship management in the hairdressing industry: an application of data mining techniques. *Expert systems with applications*, *40*(18), 7513-7518.

[60] Miguéis, V. L., Camanho, A. S., & e Cunha, J. F. (2012). Customer data mining for lifestyle segmentation. *Expert systems with applications*, *39*(10), 9359-9366.

[61] Liao, S. H., Chen, Y. J., & Lin, Y. T. (2011). Mining customer knowledge to implement online shopping and home delivery for hypermarkets. *Expert systems with applications*, *38*(4), 3982-3991.

[62] Imani, A., & Abbasi, M. (2017). Customer's clustering based on RFM model by using fuzzy c-means algorithm (case study: Zahedan City Refah Chain Store). *Management researches*, *10*(37), 251-276.