New GRFM Approach and Fuzzy LP-Metric weighting score

Elham Eshrati*, Afshin Safaee

Islamic Azad University, South Tehran Branch, Iran

ABSTRACT

Nowadays, a thorough understanding of the business processes and the organization`s customers is the essential point to survive in the market competition. In this study, a new idea was applied to import customer purchased basket data into data mining computations. First, the products were divided into families, and we assigned a numerical code for each product in the family. The sum of these assigned numbers indicates the status of the basket. After this step, the transactions were clustered based on their basket values. Customers are then clustered in each cluster using the RFM method. Using a new fuzzy LP-metric approach and pairwise comparisons, RFM indices were weighted, and we obtain customer value per cluster. Then we will proceed by averaging the customer value according to the presence of each customer in different clusters. Then we clustered customers based on customer lifetime value.

Keywords: *Data mining, RFM, LP Metric, Clustering, K-Means, Paired Comparisons, GRFM*

Introduction

Nowadays, not all customers are of equal importance to companies, and they seek to identify and analyze customer characteristics so that they can be segmented based on their value to the company (Munusamy & Murugesan, 2020; Wu & Liu, 2020). Identifying, analyzing, and segmenting customers based on their value to the company provides the basis for optimal allocation of limited resources, application of appropriate marketing strategies (Cheng & Chen,

2009), and ultimately, profitability management along with customer relationship management (Das & Mishra, 2018; Ghoreishi & Khandestani, 2019). Customer segmentation is an essential strategy for managing customer-oriented marketing efforts introduced in the 1950s to reflect a shift from mass marketing to new marketing and to targeting products or marketing operations to specific customer groups (Wong, Ton, & Wong, 2014). Customer segmentation is a supervised learning process that classifies customers into the predefined classes while customer clustering, on the other hand, groups the customers into non-predefined classes (Chang & Tsai, 2011). The discovered groups enable companies to change themselves by providing services tailored to their customers' needs and directing them to the areas where they are the most valuable customers, helping to allocate large capital, effort, and time to create the most profit (Kamthania, Pawa, & Madhavan, 2018). To identify high-response customers, the RFM model was introduced by Hughes (1994) that incorporates three variables of customer consumption interval (R value), frequency (F value), and purchase money amount (M value) to calculate customers' tendency to buy, simultaneously. The RFM model differentiates important customers from large transaction data.

To avoid any ambiguity, it should be added that the term RFM is a single value that is represented by a measurement function to integrate R, F, and M. Through RFM analysis, customer loyalty and contributions can be accurately measured (Wu & Lin, 2005). Due to the success of RFM measurement, many attempts have been made to segment the customers or cluster them based on the RFM value (Cheng & Chen, 2009; Miglautsch, 2000; Yeh, Yang, & Ting, 2008). Although RFM value has been used in customer segmentation or clustering, most previous work measures the value of RFM regardless of customers' buying behavior for different products. Therefore, the above works fail to provide effective information to promote specific products. On the other hand, given the different values of R, F, and M in different domains for experts, it can be said that there is no proper framework in which to be more accurate by industry, market, and type of customer (Chang & Tsai, 2011). What the customer buys is a good way to advertise to customers with similar shopping behavior. To discover an optimal sales policy, we need to understand potential buyers, their loyalty, and how they contribute to specific products. That is, we have to classify the customer based on the product they have purchased and calculated their RFM value to track their consumption behavior. On the other hand, in order to provide a general method that can be used in different domains and given the weight of RFM indicators, a precise method for weighting them should be developed based on the opinion of managers and experts. By doing so, we can develop detailed sales policies to meet market demand better.

Although Hughes (1994) proposed an equal weight for the three indicators of the RFM model, opinions differ on this. In some studies, the authors have used the weighted RFM model to calculate the lifetime value of the customer(Stone, 1995; Miglautsch, 2000). Some propose that the highest weight should be devoted to frequency, followed by the Recency and Monetary. Others argue that money should be the most weighted and recently acquired the least weight (Chen, Chiu, & Chang, 2005). However, the situation may vary in different industries, and managers have different opinions (Safari, Safari, & Montazer, 2016). Most of the techniques used for weighting for RFM indices have been AHP, and ANP and their fuzzy forms(Liu & Shih,

2005), which are decision-making techniques based on breaking the indices hierarchically and pairwise comparison matrix. Essentially, these methods work best when there are multiple hierarchies in the decision-making process. There is no hierarchy in the RFM problem. That is, we will only have pairwise comparisons between the three RFM indices. Based on this shortcoming in AHP and ANP techniques in the RFM case, we introduce a new fuzzy mathematical approach for weighting RFM indices.

In light of the above, in this paper, we propose a new framework for group RFM (GRFM for short) to identify loyal and highly engaged customers. GRFM calculates the customer's GRFM value, taking into account customer buying patterns as well as product features in customer analysis, rather than calculating the RFM values of customers on all goods they have purchased so far. Specifically, GRFM first discovers recurring patterns, each representing a set of goods that are repeatedly purchased in the transaction dataset. Then, based on repeated patterns discovered, customers are grouped into groups. In evaluating customer purchasing potential, we consider various product features, including average life span and unit price, and propose a new measurement function that calculates the customer's GRFM value on the products according to each recurring pattern. Therefore, we can obtain customer GRFM values in a cluster that has the characteristics of the purchased goods and accurately reflect its loyalty and participation. In addition, GRFM uses the PICC (Purchased Items-Clustering) algorithm, which can reuse discovered shopping patterns to propose appropriate sales policies to respond quickly to market demands. Also, we present a linear planning model based on the opinions of experts and industry managers based on the Lp-metric decision method for weighting the indices. The major contributions of the paper are summarized as follows:

- Create a new framework for applying the role of a customer's basket
- Calculate customer value based on the score the customer received in each cluster.
- Determine the importance of each of the RFM factors from the experts' point of view using the proposed metric LP method.
- Providing a model to predict customer behavior and value according to the customer's basic attributes.

The rest of this paper is presented as follows. In Section 2, we review related works and previous literature. In section 3, we introduce the new GRFM framework. Section 4 describes our empirical results, and finally, Section 6 contains discussions and conclusions.

Literature review
CRM
Today, new business culture is emerging in which customer-based economies are fundamentally changing, and companies are looking for solutions that address these challenges. Manufacturing companies today face increasing customer value throughout the life cycle of each customer purchasing from the company. Moreover, the old design-build-resell paradigm has become the new resell-build-redesign paradigm (converting product-oriented to customer-oriented)(Parvatiyar & Sheth, 2001). In the customer-centric domain, one of the most important factors in customer

loyalty to a business is the concern of companies today to build a loyal customer base. This loyalty is important both in terms of encouraging future purchases and as an indicator of how customers are communicating with their customers (Lam, Shankar, Erramilli, & Murthy, 2004). By equipping companies with customer loyalty measurement tools, a supplier will be able to understand how its efforts to maintain good relationships can help its level of profit. Different authors have put forward different theories to measure the amount of customer loyalty in marketing. For example, it has been shown that a direct relationship with the customer has a direct effect on customer loyalty (Richard & Perrien, 1999). Other researchers examined the effect of other constructs such as customer satisfaction, quality of customer relationships, trust, partnerships, and organizational change on customer loyalty (Chow & Holden, 1997). Customer Relationship Management (CRM) is a customer-oriented approach. CRM focuses on long-term relationships with customers by providing the interests and values of the customer (not what the company is looking for). CRM attempts to answer two main questions (Gary & Boon, 2001):

 -What is the benefit to the customer?

- How to increase customer value?

In other words, CRM is considered as a business operation philosophy aiming to attract and maintain customers, increase their value and loyalty, as well as implement customer-oriented approaches (Nenonen & Storbacka, 2015).

CLV

In the marketing literature, in order to have an effective CRM, companies must adopt customer-centric approaches such as customer value assessment (Kotler, 2009). Customer value is the value that a customer gives to a company over its lifetime. In evaluating customer value, companies must consider which criteria are important in identifying customer value. The concept of CLV (customer lifetime value) in the CRM context is considered an appropriate benchmark for marketing in evaluating customer value (Kumar, 2008). The CLV was defined in the mid-1970s as "the present value of expected future earnings flows over a customer time horizon" (Kotler, 1974). In recent years, CLV and its applications have received more attention(Reinartz & Kumar, 2000). Some of these studies have proposed models that calculate CLV using past customer data, while others consider future behavior due to a lack of past data (Grover & Vriens, 2006). The effect of customer purchasing behavior on the future profitability of a company can be shown by CLV (Chang et al., 2012). The CLV approach combines two marketing and financial contexts, thereby enabling management and optimization (C. Williams & R. Williams, 2015). The main applications of CLV are business-to-consumer (B2C) applications, and business-to-business (B2B) applications of customer asset management are neglected (Nenonen & Storbacka, 2015). The relevance of CLV applications is primarily used at the customer level, which affects customer loyalty (Qi, Qu, Zhou, & Li, 2015) and maintains the influence of customer behavior. According to the results (Kumar & Pansari, 2016), the economy of the country has a direct impact on the frequency of purchases, and the margin of purchases and attention to cultural dimensions and economic conditions should be considered due to their different impact on CLV. All of these (and other) elements with the given model and set directly or indirectly affect the calculation of the CLV.

Customer segmentation

Customer segmentation is one of the most effective ways to manage different customers with different preferences. In this process, heterogeneous customer groups are divided into homogeneous groups based on similar features and differences in dissimilar features (Quinn, Hines, & Bennison, 2007). Dividing customers into homogeneous groups increases both customer satisfaction and the expected profit of a company. Applying different marketing strategies to different customer groups can increase customer value and enable companies to plan long-term relationships with customers by meeting customer needs(Quinn, Hines, & Bennison, 2007). Also, companies can improve their income by achieving and retaining valuable customers. Based on (Dursun & Caber, 2016)among segmentation methods, clustering is one of the most useful methods for identifying homogeneous groups of customers and formulating customized marketing strategies for each group (Liu & Shih, 2005). One of the most common ways to clustering customers is to segment them based on the scores they receive in the RFM model, which measures customer loyalty and customer lifetime value (CLV). The use of the RFM model is to calculate the CLV with an emphasis on the profitability of each customer (Hu & Yeh, 2014). But given that each customer has a different purchase basket, the use of the RFM model in general for all customers does not consider the role of customer preferences in purchasing different goods. To address this problem, Chang and Tsai (2011) proposed a group RFM method. In their proposed method, each commodity is first assigned by an integer. The customer basket purchased by each customer consists of a number of these items, which are now assigned by a number. The sum of these numbers in each basket indicates it has purchased status. These numbers will be largely close to baskets. Operationally in application situations, their proposed algorithm is difficult to use and requires programming, and it is practically impossible to use commercial software such as SPSS Modeler and RapidMiner.

K-Means algorithm

Clustering is the process of grouping objects into similar groups that we call them to cluster. A cluster is a set of data objects that are similar to one another in a cluster and are not similar to those in other clusters (Han &Kamber, 2001). K-Means is one of the well-known algorithms for clustering that is very sensitive to the choice of starting point for partitioning cases in early K clusters. One can compare the performance of different clustering methods using the intraclass method when the number of fixed cluster of K value is defined as $F(K) = 1/k \sum_{n=1}^{k} \sum_{c_i} Dist(c_i, c^n)$ (Shina & Sohnb, 2004). Given the above considerations, we seek to choose a suitable method for clustering customers based on the purchased basket and their CLV. To this aim, this paper provides a proper algorithm by combining the lp-fuzzy optimization and clustering algorithm.

Methodology for New GRFM framework

The proposed new GRFM framework is an extension for the proposed method by Chang and Tsai (2011), which has three phases and initiates with two initial steps in the first phase. This phase includes classifying purchased items based on their properties and assigning a number to each customer basket (constructing the Original Patterns (ORPA) table). The notations used are listed in Table (1).

Table 1. Summary of the notations utilized in this paper.

D	*Transaction dataset*
M	*Categorical data records*
d_i	*Some data items*
I	*data item set that contains all the data items in D*
N	*Number of items*
$Item_i$	*i th stuff*
SI	*Subset of I*
$VAL(Item_i)$	*Integer allocator function for stuffs*
$F_T(d_i)$	*Conversion function that used to convert i th to the integer data D*
IP_i	*i th data in ORPA*
$IP - num_i$	*Number of IP_i*

In order to assign a number to purchased items, they must be first classified according to their properties into k categories and then set a unique binary number to every purchased item with the following function:

(1)
$$VAL(Item_i) = 2^{i-1}, i \text{ is an index}$$

For each basket, the sum of the scores ($F_T(d_i)$) obtained from the above function is calculated.

(2)
$$F_T(d_i) = \sum_{j=1} VAL(Item_{i,j})$$

As discussed earlier, the contents of the customer baskets (purchased items in a customer transaction) are of great importance and represent their desires. In our new GRFM framework, the second phase is the clustering of the baskets scores. Clustering the scores obtained from the first phase puts customers together in each cluster whose basket score is close together. As such, clusters are formed whose behavior and buyers' preferences in each cluster are almost identical. If the baskets are the same, the scores are equal, and they must be in the same cluster; in the case of some commodities, the totals in the baskets will be somewhat variable, and in the case of proximity, they will still be in the same cluster. At the clustering phase, the number of desirable clusters must first be determined for the obtained data from the previous phase. For this purpose, two indices of Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) are used. The two criteria of Bayesian and Akaike show how much use of a statistical model causes

the loss of information. In other words, these criteria strike a balance between model accuracy and complexity. The BIC is commonly defined as:

(3) $$BIC = ln(n)k - 2\,ln\,(\hat{L})$$

Where L is the maximum value of the likelihood function for a model such as M, $\hat{L} = p(x|\hat{\theta}, M)$, x observed data, n is the number of data and k is the number of estimated parameters. Each cluster includes customers who have a similar purchased basket and are comparable to each other because they have close IPs.

The clusters formed in the previous phase, which contain similar purchased baskets, are grouped in the third phase based on RFM indices. In other words, each customer in each cluster will have their unique (R, F, M) values based on the items in their purchased baskets. The presented framework does not calculate only one (R, F, M) value for each customer, but different (R, F, M) values for each customer based on the presence of the customer in different clusters. Customers' behavior may vary in purchasing different items; some are purchased more, and others are purchased at larger intervals and less frequently, with different financial amounts.

At this point, the status of the clusters, the groups within the clusters, the RFM factors in each cluster, and the degree of importance given to these factors by experts, are critical to determining the value of each customer (CLV) for the company. To calculate customer value (CLV), we first need to determine the weight of the coefficients of R, F, and M based on experts' opinions. For this purpose, we have presented a new approach which will be explained in detail in the next section.

To analyze the status of each customer in each group in the clusters and before calculating the CLV, the values of the RFM factors were standardized with the help of the following formula. Because the higher the factors than the minimum, the better for the customer, the following formula was used for a client i in my cluster j:

(4) $$x_{i,j} = \frac{x_{i,j} - x_{min}}{x_{max} - x_{min}}$$

Each customer in each group in the clusters has a CLV. By averaging between the CLVs of each customer, the final value of each customer is determined. It is now possible to classify customers accordingly.

Weighting RFM Indicators

In the RFM model, the value of customer C_i is measured by eq. (4).

$$V(C_i) = W^R R(C_i) + W^F F(C_i) + W^M M(C_i) \qquad (4)$$

Where $R(c_i)$, $F(C_i)$ and $M(C_i)$ show each customer (C_i) R, F and M values and W^R, W^F and W^M represent their weight, respectively. Because of the different weightings of the indicators for different experts and managers, we need first to determine the weight of each one. To determine

the weight of the RFM indices, the pairwise matrix with fuzzy numbers is used. This paper proposes a new method for the evaluation of weights. In this method, we will solve an optimization model based on the Lp-metric method with the infinite norm. In general, the Lp-metric method solve the following model with assuming $F_1, F_2 \ldots, F_k$ functions and their ideals $F_1^* \ldots, F_k^*$ by considering p between 1 and ∞ :

$$Min \; L-P = \left(\sum\nolimits_{i=1}^{k} \left(w_i (F_i^* - F_i)^p \right) \right)^{\frac{1}{p}} \tag{2}$$

S.t:

$$X \in S$$

Although the ideal point for all functions in the Lp-metric model is achieved rarely, it does achieve the point that ultimately creates the minimum distance to the ideal. In this problem, S is the feasible region, and with increasing p, Lp problem changes to finding the minimum value of maximum distances between each function and their ideals, so:

$$L - \infty = Min \; \lambda \tag{3}$$

S.t:

$$\Box \geq (F_i^* - F_i) \qquad \forall i = 1, 2 \ldots, k$$

$$\Box \geq 0$$

The result of the pairwise matrix after defuzzification is a matrix-like below:

$$D = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

If the matrix is 2 x 2, there is no conflict between the comparisons. Conflict occurs when the dimensions of the matrix are larger and different people make these comparisons. Each element of the pairwise comparisons matrix is the weight distribution that the expert compares to each case. As a result, the weights of the indices are calculated as $W_1 \ldots, W_k$ so that the ratio W_i / W_j is even as close as possible to a_{ij}. To do this, we base the $L - \infty$ solution on the following model:

$$Min \; \lambda$$

S.t:

$$\Box \geq Max \left\{ \left\| \frac{W_i}{W_j} - a_{ij} \right\| \right\} \qquad i = 1, 2 \ldots n \quad .j = 1, 2 \ldots, m$$

$$\sum_{i=1}^{k} W_i = 1$$

$$\lambda \geq 0$$

$$W_i \geq 0 \, , i = 1, 2 \ldots, k$$

Experiments

Data

In our experiment, the samples of purchase data were obtained from Pioven Trading Company (Iran), which distributes cosmetics, between 2018 and 2019. The business model of this company is B2B (business to business), and its customers are of 6 types include hairdressers, pharmacies, dentists, chain stores, supermarkets, and small shops. The biggest number of customers are small shops, and the smallest number is hairdressers. The variety of customer purchased basket is classified into three categories: low, medium, and high. We found a relationship between the type of customers and the variety of the basket. The variety of baskets in small shops are medium to large. In contrast, in the supermarkets and hairdressers, the variety of the purchased baskets is low, medium, and high, which is distributed almost equally. In the case of pharmacies and dentists, a variety of purchased baskets is almost medium.

Implementation of Group RFM Model

Using CRISP as a data mining process basic model, we first identified the business and the data. After this step, given our mastery over the data behaviors, we entered the third phase of this process, data preparation. As detailed in the previous section, our new GRFM framework has three main phases. In the first phase, we classified the items based on their properties and allocated numbers to each and calculated the score for the customers purchased baskets (ORPA table). In analyzing customer transactions during the research period, it was found that 50 goods were purchased in two cosmetics and sanitary categories and from different brands. To assign a binary number to each transaction, the product classification was performed first. For each item ($Item_i$), each transaction was assigned an IP. If one of the items was present, the corresponding registration number (2^{i-1}), and finally, with the sum of the values assigned, the IP value was calculated. Next, based on the IP values obtained, clustering was performed by the K-means method on scores of baskets.

Table 2. Classification and integer assigned to each item

Product classes	Brand	Product	i	Allocated Integer
cosmetics	Brand 1 she	I1	0	1
		I2	1	2
		I3	2	4
		I4	3	8
	Brand 2 brush	I5	4	16
		I6	5	32
		I7	6	64
		I8	7	128
	Brand 3 caldion	I9	8	256
		I10	9	512
		I11	10	1024
		I12	11	2048
	Brand 4	I13	12	4096

	rebul	*I14*	*13*	*8192*
		I15	*14*	*16384*
		I16	*15*	*32768*
	Brand 5 *desert*	*I17*	*16*	*65536*
		I18	*17*	*131072*
		I19	*18*	*262144*
		I20	*19*	*524288*
	Brand 6 *odora*	*I21*	*20*	*1048576*
		I22	*21*	*2097152*
		I23	*22*	*4194304*
		I24	*23*	*8388608*
	Brand 7 *valera*	*I25*	*24*	*16777216*
		I26	*25*	*33554432*
		I27	*26*	*67108864*
		I28	*27*	*134217728*
	Brand 8 *aromel*	*I29*	*28*	*268435456*
		I30	*29*	*536870912*
		I31	*30*	*1073741824*
		I32	*31*	*2147483648*
	Brand 9 *fonex*	*I33*	*32*	*4294967296*
		I34	*33*	*8589934592*
		I35	*34*	*17179869184*
		I36	*35*	*34359738368*
sanitary	*Brand 1* *Trisa electric*	*I37*	*36*	*68719476736*
		I38	*37*	*1.37439E+11*
		I39	*38*	*2.74878E+11*
		I40	*39*	*5.49756E+11*
	Brand 2 *toothbrush*	*I41*	*40*	*1.09951E+12*
		I42	*41*	*2.19902E+12*
		I43	*42*	*4.39805E+12*
	Brand 3 *Dental floss*	*I44*	*43*	*8.79609E+12*
		I45	*44*	*1.75922E+13*
		I46	*45*	*3.51844E+13*
		I47	*46*	*7.03687E+13*
	Brand 4 *Tooth paste*	*I48*	*47*	*1.40737E+14*
		I49	*48*	*2.81475E+14*
		I50	*49*	*5.6295E+14*

Clustering transactions by the K-means method

The number of desirable clusters must first be determined for the data. For this purpose, two indices of Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) were used. Accordingly, by applying these variables, 10 clusters were selected for the best number of clusters for data. The model summary is shown in Figure (1).
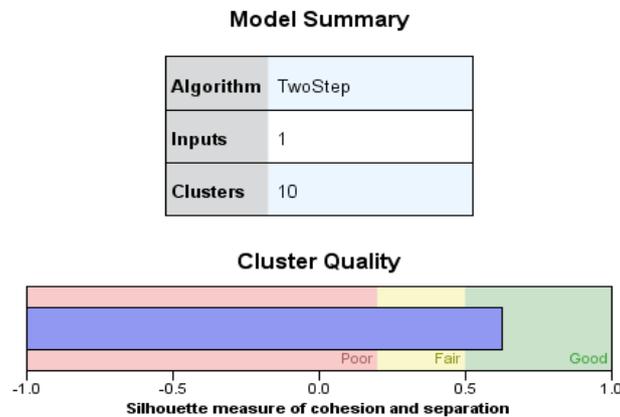
Model Summary

Algorithm	TwoStep
Inputs	1
Clusters	10

Cluster Quality

Figure 1. Summary of the clustering model

Table (3) shows the center of each cluster. Each cluster includes customers who have a similar product portfolio and are comparable to each other because they have close IPs. Grouping after this step will make shopping carts more similar. The result of this type of analysis is the comparison of customers based on the type of portfolio along with the RFM factors.

					Cluster					
10	9	8	7	6	5	4	3	2	1	
2.E+014	5.E+013	1.E+015	7.E+014	2.E+014	4.E+014	9.E+013	8.E+014	5.E+014	1.E+013	*Numeric IP value*

Table 3. Location of cluster centers

Next, the status of M, R, and F is determined for each cluster. Each customer can fit into different clusters. The calculations made up to this point identified several factors. The purpose of clustering based on items in baskets gives this feature to the framework used in this study to compare similar baskets and to extract RFM factors. This means that customers' behavior may vary in purchasing different items; that is, some are purchased more, and others are purchased at larger intervals and less frequently and with different financial amounts. In other words, customers in different clusters will have different RFM factors. At this point, the status of the clusters, the groups within the clusters, the RFM factors in each cluster, and the degree of importance given to these factors by experts, are critical to determining the value of each customer (CLV) for the company. Table (4) shows the properties of groups within each cluster. To analyze the status of each customer in each group of clusters and before calculating the CLV, the values of the RFM factors were standardized.

*Table 4.*Properties of groups

Cluster	Group	Avg. purchase amount	Avg. ofR	Avg. of F	Avg. of M	Number of cluster member
1	1	2354.7202	1.62	2.99	2.96	390
	2	3482.5706	4.05	3.05	3.05	506
2	1	16074.0714	1.49	3.01	3.01	394
	2	17815.1082	4.00	2.98	3.00	606
3	1	20739.1122	3.49	2.69	3.02	228
	2	22375.9206	3.54	3.47	3.00	167
	3	20238.1933	2.00	2.99	3.00	184
	4	19506.9256	1.00	3.00	3.01	209
	5	21131.3570	5.00	3.01	3.00	212
4	1	6389.6141	4.50	3.00	3.02	399
	2	4953.6477	2.00	3.03	3.00	590
5	1	14673.0159	3.98	2.96	3.02	610
	2	12848.2493	1.49	2.97	3.01	390
6	1	8845.9753	4.53	2.99	3.01	406
	2	7376.9351	1.99	3.04	3.01	591
7	1	19880.6496	3.55	3.03	3.00	204
	2	19298.1191	1.48	2.95	3.00	397
	3	18059.0732	1.00	3.06	3.01	399
8	1	21267.1710	1.50	3.01	3.01	387
	2	23527.1706	3.91	3.00	3.01	613
9	1	2359.4783	1.00	2.79	3.02	194
	2	4229.5695	3.00	2.67	3.01	189
	3	3658.9345	2.00	3.06	3.01	192
	4	4295.5778	4.00	2.73	3.01	197
	5	4783.0832	5.00	2.94	3.03	191
10	1	11909.7559	3.49	2.97	3.01	398
	2	11920.4543	5.00	3.03	3.00	198
	3	10263.6324	2.00	2.95	3.02	203
	4	9426.1483	1.00	2.98	3.00	200

Customer Lifetime Value Calculation (CLV)

At this point, we asked managers to determine the importance of each of the FRM factors to customers by pairwise comparison between them. In this section, the customer lifetime value (CLV) is calculated for each group. This CLV is obtained based on an integrated rate for each group. The integrated rate of group j is:

$$C_i^j = w_R C_R^j + w_F C_F^j + w_M C_M^j$$

Where C_R^j, C_F^j and C_M^j are the average of the normalized RFM values for customers in each group and w_R, w_F and w_M are the weights of the indices calculated by the method presented in section 3.1. The following steps were performed to calculate weights:

Step 1. Company managers were asked to compare the importance of the four indicators of the RFM model with the length of the relationship with the company.

Step 2. Managers' statements become triangular fuzzy numbers.

Step 3. The fuzzy averaging between the views was calculated, and the fuzzy operation was performed. Table (5) shows the final result of the first three steps.

Table 5. The result of the first three steps of weighting the indicators

Paired comparisons	Length of contact with the company	Recency (R)	Frequency (F)	Monetary (M)
Length of contact with thecompany	1	2.787	3.354	2.208
Recency (R)	0.359	1	0.53	0.336
Frequency (F)	0.283	1.888	1	1.765
Monetary (M)	1.888	2.98	0.566	1

Step 4. The proposed L-P model was used for weight calculation in Lingo 14.0. The values of W1, W2, W3, and W4 were calculated as 0.41388, 0.1339, 0.2420, and 0.2100, respectively. It seems to the managers of the company that the length of customer relationships is more important than other indicators. Given the above values and the normal values, the following four indices are used to calculate the CLV:

$$C_i^j = 0.1339\ C_R^j + 0.2420 C_F^j + 0.2100 C_M^j$$

Then averaging was performed between the CLVs of each customer. After calculating the value of each customer, it is now possible to classify the customer accordingly. Based on the AIC criteria, the best number of proposed clusters was selected for these values of 8. Table (6) shows the values of this factor, and Figure (2) shows the variations of this factor per number of clusters. The status of RFM indices in each of the clusters is given in Table (7).

Table 6. AIC factor changes per number of clusters

Number of clusters	Akaike's Information Criterion (AIC)	AIC Change[a]	Ratio of AIC Changes[b]	Ratio of Distance Measures[c]
1	*2089.941*			
2	*1781.789*	*-308.152*	*1.000*	*1.350*
3	*1556.668*	*-225.121*	*.731*	*1.284*
4	*1384.045*	*-172.623*	*.560*	*1.272*
5	*1250.897*	*-133.148*	*.432*	*1.769*
6	*1180.862*	*-70.035*	*.227*	*1.112*
7	*1119.059*	*-61.803*	*.201*	*1.026*
8	*1059.121*	*-59.938*	*.195*	*1.579*
9	*1025.565*	*-33.556*	*.109*	*1.090*
10	*995.786*	*-29.779*	*.097*	*1.022*
11	*966.889*	*-28.897*	*.094*	*1.114*
12	*942.188*	*-24.700*	*.080*	*1.061*
13	*919.587*	*-22.601*	*.073*	*1.101*
14	*900.162*	*-19.425*	*.063*	*1.067*
15	*882.722*	*-17.440*	*.057*	*1.003*
16	*865.382*	*-17.340*	*.056*	*1.117*
17	*851.126*	*-14.256*	*.046*	*1.051*
18	*838.151*	*-12.975*	*.042*	*1.348*
19	*831.619*	*-6.532*	*.021*	*1.011*
20	*825.292*	*-6.327*	*.021*	*1.123*

Auto-Clustering
Akaike's Information Criterion (AIC)

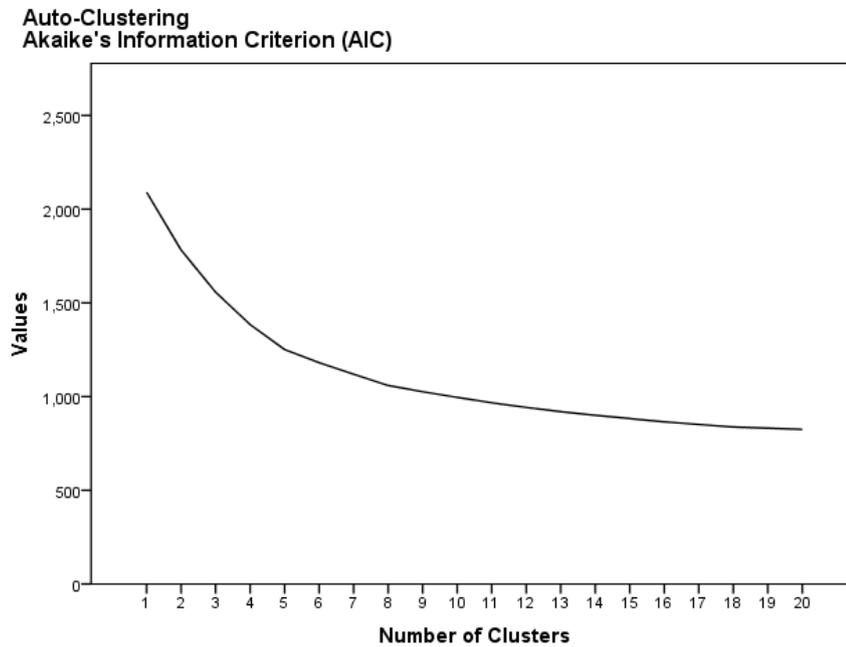*Figure 2.* Changes in AIC factor to increase the number of clusters

Table 7. Status of RFM factors in formed clusters

		mean R		mean F		mean M	
		Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
Clusters	1	3.3915	.29704	3.4003	.31805	3.3020	.26909
	2	3.2706	.32717	3.5177	.28105	2.6000	.24725
	3	2.3988	.34527	3.6446	.24862	3.0018	.38707
	4	2.4910	.25200	2.8822	.27833	2.8960	.25836
	5	2.7939	.40553	2.9420	.30524	3.7685	.23855
	6	3.0533	.28835	2.6509	.35756	2.4492	.29741
	7	3.7916	.23312	2.5994	.35344	2.9191	.30775
	8	3.0811	.26377	2.4217	.28969	3.3089	.26018
	Combined	3.0064	.51357	2.9850	.51091	3.0076	.48898

Investigating customer attributes and naming on clusters

After analyzing the RFM factors in the eight final clusters, we calculated the eight final clusters to calculate the weighted value. The clusters were assigned the names of Gold plus, Gold, Silver Plus, Silver, Bronze Plus, Bronze, Platinum Plus, Platinum based on RFM values.

Conclusion

This research was conducted in the framework of the CRISP data mining method and presented a new framework. The thing that is overlooked in the conventional RFM method is the product basket that the customer orders each time. A customer purchased a basket with a higher financial value or, like other customers, with a higher frequency of purchase than a customer purchased a basket of less value or high value but with low multiplicity, should be considered differently. In this study, a new idea was applied to import customer purchased basket data into data mining computations. First, the products were divided into families, and we assigned a numerical code for each product in the family. The sum of these assigned numbers indicates the status of the basket. We named this value IP. Each customer purchased basket will have an IP, and if the customer purchased basket is the same, they would have the same IP value. After this step, the transactions were clustered based on IP values. The baskets of similar goods fall into a cluster. The RFM model for each of the clusters was performed in SPSS 24 software. In order to arrange the transactions more precisely within these clusters, we classified the groups within each cluster based on recent exchange data (R), the number of transactions (F), and monetary exchange value (M).

Customers in each cluster find this grouping more precisely. In other words, each customer in each cluster will have their RFM values based on the baskets they have purchased. By normalizing the RFM factors in each group, these values were calculated for each customer and the customer lifetime value (CLV) was calculated and finally averaged between the CLVs of each customer to have a final CLV. Here are two main features of the framework outlined below: the framework presented does not only calculate a value (R, F, M) for a customer but assigns different values (R, F, M) to each customer purchased basket characteristics. This will determine the correct customer behavior. Besides, this framework specifies a cube (R, F, M) for each cluster, indicating the client's position in that cluster. These cubes (R, F, M) not only support the traditional RFM method based on Miglatash's (2000) model but also allow new managers to analyze. Based on this information, the manager can properly contact customers and plan more personal purchases. Secondly, the provided framework provides sales information for each cluster purchases that are clustered according to the properties of the purchased items. Based on the information, the user can obtain integrated sales information, for example, the buying cluster is highly loyal and profitable, or the purchased product has a large potential sales volume. This information can be reviewed by the administrator to analyze important hidden information. Therefore, the manager can base this information on establishing an optimal inventory management system to reduce the risk of excess stock. Note that this information cannot be obtained from the usual RFM analysis paradigms.

References

Chang, H.-C., & Tsai, H.-P. (2011). Group RFM analysis as a novel framework to discover better customer consumption behavior. *Expert Systems with Applications, 38*, 14499-14513.

Chen, C., Chiu, A., & Chang, H.-H. (2005). Mining changes in customer behavior in retail marketing. *Expert Systems with Applications*, 773-781.

Cheng, C. H., & Chen, Y. S. (2009). Classifying the segmentation of customer value via rfm model and rs theory. *Expert System with Application, 36*(3), 4176-4184.

Chow, S., & Holden, R. (1997). Toward an understanding of loyalty: The moderating role of trust. *Journal of Management*, 275-290.

Das, S., & Mishra, M. (2018). *The Impact of Customer Relationship Management (CRM) Practices on Customer Satisfaction.* Rajagopal: Behl R. (eds) Business Governance and Society.

Dursun, A., & Caber, M. (2016). Using datamining techniques for profiling profitable hotel customers: An application of RFM analysis. *Tourism Management Perspectives*, 153-160.

Ghoreishi, S., & Khandestani, K. (2019). Customer Segmentation Based on GRFM: Case Study. *World of Computer Science and Information Technology Journal, 9*(1), 1-6.

Grover, R., & Vriens, M. (2006). *The handbook of marketing research: Uses, misuses, and future advances.* Sage Publications, Inc.

Hu, Y.-H., & Yeh, T.-W. (2014). Discovering valuable frequent patterns based on RFM analysis without customer identification information. *Computer Science*, 76-88.

Hughes, A. (1994). *Strategic database marketing. Chicago.* Probus Publishing Company.

Kamthania, D., Pawa, A., & Madhavan, S. (2018). Market Segmentation Analysis and Visualization using K-Mode Clustering Algorithm for E-Commerce Business. *Journal of Computing and Information Technology*, 57-68.

Kotler, P. (2009). *Marketing Management.* New Jersey: Prentice Hall.

Kumar, V., & Pansari, A. (2016). Competitive Advantage through Engagement. *Journal of Marketing Research*, 50-62.

Lam, S., Shankar, V., Erramilli, M., & Murthy, B. (2004). Customer value, satisfaction, loyalty, and switching costs: An illustration from a business-to business service context. *Journal of the Academy of Marketing Science*, 293-301.

Liu, D., & Shih, Y. (2005). Integrating AHP and data mining for product recommendation based on customer lifetime value. *Information & Management*, 387-400.

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

Miglautsch, J. (2000). Thoughts on rfm scoring. *Journal of Database Marketing*, 67-72.

Munusamy, S., & Murugesan, P. (2020). Modified dynamic fuzzy c-means clustering algorithm – Application in dynamic customer segmentation. *Applied Intelligence, 50*, 1922-1942.

Nenonen, S., & Storbacka, K. (2015). Driving shareholder value with customer asset management: moving beyond customer lifetime value. *Industrial Marketing Management*, 140-150.

Parvatiyar, A., & Sheth, J. (2001). Customer relationship management: Emerging practice, process and discipline. *Journal of Economic and Social Research*, 1-34.

Qi, J., Qu, Q., Zhou, Y., & Li, L. (2015). The impact of users' characteristics on customer lifetime value raising: Evidence from mobile data service in China. *Inf. Technol. Manag*, 273-290.

Quinn, L., Hines, T., & Bennison, D. (2007). Making sense of market segmentation: a fashion retailing case . *European Journal of Marketing*, 439-465.

Reinartz, W., & Kumar, V. (2000). *On the profitability of long lifetime customers: An emprical investigation and implication for marketing.* Fontainebleau, France.

Richard, L., & Perrien, J. (1999). Explaining and evaluating the implementation of organizational relationship marketing in the banking industry: Clients' perception. *Journal of Business Research*, 199-209.

Safari, F., Safari, N., & Montazer, A. G. (2016). Customer lifetime value determination based on RFM model. *Marketing Intelligence & Planning*, 1129-1157.

Shina, H., & Sohnb, S. (2004). Segmentation of stock trading customers according to potential value. *Expert Systems with Applications*, 27-33.

Stone, B. (1995). *Successful direct marketing methods.* Lincolnwood: NTC Business Books.

Wong, R., Ton, C., & Wong, A. (2014). Examine the Effects of Customer Satisfaction on Customer Loyalty: An Empirical Study in the Healthcare Insurance Industry in Hong Kong. *British Journal of Economics, Management & Trade, 4*(3), 372-399.

Wu, J., & Lin, Z. (2005). Research on customer segmentation model by clustering. *Proceedings of the 7th ACM ICEC international conference on electronic commerce.*

Wu, T., & Liu, X. (2020). A dynamic interval type-2 fuzzy customer segmentation model and its application in E-commerce. *Applied Soft Computing, 9.*

Yeh, C. I., Yang, K.-J., & Ting, M. T. (2008). Knowledge discovery on rfm model using bernoulli sequence. *Expert Systems with Application*, 5866-5871.

Ethics Declarations

Competing Interests
No, there are no conflicting interests.

Rights and Permissions