# Data-mining application for country segmentation based on the RFM model

## Mehdi Ghazanfari

Industrial Engineering Department
Iran University of Science and Technology
Tehran, Iran
E-mail: mehdi@iust.ac.ir

## Samira Malek Mohamadi*

Industrial Engineering Department
Iran University of Science and Technology
Tehran, Iran
Fax: +98–21–22244181
E-mail: samira.malek@gmail.com
*Corresponding author

## Somayeh Alizadeh

Industrial Engineering Department
Iran University of Science and Technology
Tehran, Iran
E-mail: s_alizade@mail.iust.ac.ir

**Abstract:** For effective Customer Relationship Management (CRM), it is important to gather information on customer value. Segmentation is the method of knowing the customers and partitioning a population of customers into smaller groups. This paper develops a novel country segmentation methodology based on Recency (R), Frequency (F) and Monetary value (M) variables. After the variables are calculated, clustering methods (K-means and fuzzy K-means) are used to segment countries and compare the results of these methods by three different criteria. Customers are classified into four tiers: Top-active, Medium-active, New customer and Inactive. Then a customer pyramid is drawn and the customer value is calculated. Consequently, the data are used to analyse the relative profitability of each customer cluster and the proper strategy is determined for them.

**Biographical notes:** Mehdi Ghazanfari received his BS degree in Industrial Engineering from the Iran University of Science and Technology, Tehran, in 1983, his MS degree in Industrial Engineering from Amir Kabir University, Tehran, Iran, in 1986 and his PhD from the University of NSW, Sydney, Australia, in 1996. He is a Professor at the Department of Industrial Engineering of the Iran University of Science and Technology from 1997 up to now.

Samira Malek Mohamadi received her BS degree in Chemical Engineering from the University of Science and Research, Tehran, in 2003 and her MS degree in Information Technology Engineering from the Iran University of Science and Technology, Tehran, Iran, in 2006.

Somayeh Alizadeh received her BS degree in Computer Engineering from Sharif University of Technology, Tehran, Iran, in 1997, her MS degree in Industrial Engineering from the Iran University of Science and Technology, Tehran, in 2002 and her PhD in Industrial Engineering from the Iran University of Science and Technology, Tehran, in 2008.

# 1 Introduction

The mass marketing approach cannot satisfy the needs of today's diverse customers. This diversity should be satisfied using segmentation that divides markets into customer clusters with similar needs and/or characteristics that are likely to exhibit similar purchasing behaviours (Dibb and Simkin, 1996). The concept of segmentation is central to Customer Relationship Management (CRM). Segmentation means partitioning a population of customers into different segments, considering the most within-segment homogeneity and between-segment heterogeneity. Segmentation is valuable because it allows the end user to look at the entire database from a much higher level. It also allows a company to differentially treat consumers in different segments. One-to-one marketing is the ideal marketing strategy, in which every marketing campaign or product is optimally targeted at each individual customer; but this is not always possible. Thus, segmentation is required to distinguish similar clients and put them together in a segment. Doubtlessly, using segmentation to understand customers' needs is much easier, faster and more economical than uniquely investing to understand them particularly (Berson *et al*., 2001).

There are different methods for country segmentation. Helsen *et al*. (1993) developed an approach to country segmentation with multinational diffusion patterns for three different consumer durable goods and modified the Bass diffusion model to determine the segments and segment level estimates of diffusion patterns. Tsai and Chiu (2004) developed a market segmentation methodology based on product-specific variables such as purchased items and the associative monetary expenses from the transactional history of customers to address the unreliable results of segmentation based on general variables like customer demographics. Shina and Sohnb (2004) used three clustering methods (K-means, the self-organising map and fuzzy K-means) for segmentation to find properly graded stock market brokerage commission rates based on transactional data. Jonkera *et al*. (2004) presented an approach to segment customers based on Recency (R), Frequency (F), and Monetary value (M) variables. Hwang *et al*. (2004)

introduced a framework for analysing customer value and segmenting customers based on their current value, potential value and customer loyalty. Golsefid *et al.* (2007) segmented countries according to the analysed countries' export baskets and the correlation between groups' commodities, and defined Dissimilarity Export Basket (DEB) functions to measure the similarity among the countries.

The main role of the Trade Promotion Organization (TPO) of Iran is determining policy and strategy in foreign trade. In this study, we segment the customers (countries) of the TPO. In fact, each country is assumed to be a customer and the TPO tries to segment countries into homogeneous groups based on the RFM model variables, which are important parameters in foreign trade. RFM indicate export stability, consequently maintaining recent transactions, volume and profitability. According to the segmentation results, the TPO customer pyramid is drawn and the proper strategies are determined to keep, retain and raise the value for each customer group.

First we segment countries using clustering algorithms and the RFM variables. Later, in order to suggest appropriate strategies for promoting each segment, we analyse each cluster using a customer pyramid, a decision tree and the RFM model.

The remainder of this paper is organised as follows: Section 2 examines the framework of the proposed country segmentation methodology, including data preparation, calculating R, F and M amounts, clustering algorithms and evaluating the clustering quality. In Section 3, we analyse each cluster value and profitability using a decision tree, a customer pyramid and the RFM model. In Section 4, the proposed country segmentation methodology is put into practice using the TPO database. Finally, in Section 5, the implication of the results is discussed and further study areas are suggested.

**Figure 1**   Framework of country segmentation based on R, F and M
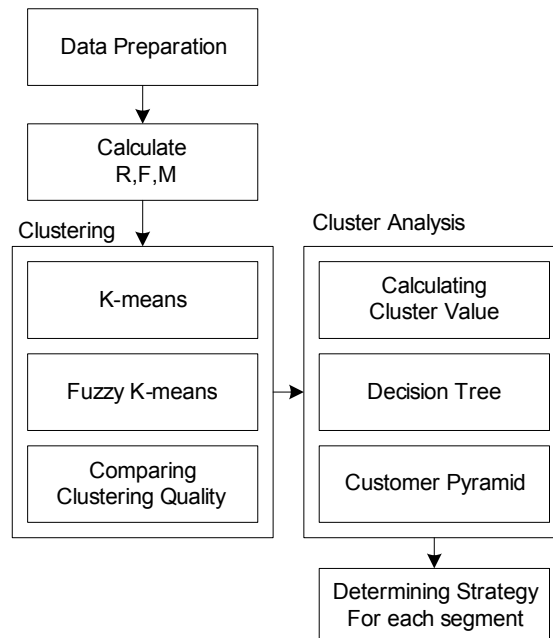
## 2    Country segmentation based on R, F and M

This section introduces a novel country segmentation methodology based on the RFM variables. The core of the methodology includes data preparation, calculating the RFM amounts, clustering by K-means and fuzzy K-means algorithms, and comparing the quality of these clustering methods.

### 2.1    Calculating the RFM variables

After data preparation, which is the first essential part of the data-mining procedure, we calculated the RFM criteria. 'Recency' measures the interval between the most recent time we had export to each of the countries and the analysing time. 'Frequency' measures the export frequency within a specified period. 'Monetary' measures the total monetary value within a specified period.

The scores can vary depending on the types of applications and scoring approaches (Jonkera *et al.*, 2004). The scores retrieved from the original transaction database are normalised for clustering purposes. Therefore, the $R(x_i)$, $F(x_i)$ and $M(x_i)$ scores can be redefined as follows:

$$R(x_i) = \frac{Q^R - Q_{Min}^R}{Q_{Max}^R - Q_{Min}^R} \tag{1}$$

$$F(x_i) = \frac{Q^F - Q_{Min}^F}{Q_{Max}^F - Q_{Min}^F} \tag{2}$$

$$M(x_i) = \frac{Q^M - Q_{Min}^M}{Q_{Max}^M - Q_{Min}^M} \tag{3}$$

where

$Q^R$, $Q^F$ and $Q^M$ = the original values for a country $x_i$ according to the definitions of R, F and M

$Q_{Min}^R$, $Q_{Min}^F$ and $Q_{Min}^M$ = the minimum values of R, F and M

$Q_{Max}^R$, $Q_{Max}^F$ and $Q_{Max}^M$ = the maximum values of the same.

### 2.2    Clustering

The process of grouping a set of objects into classes of similar objects is called clustering. A cluster is a collection of data objects that are similar to one another and are dissimilar to the objects in other clusters (Ye, 2003). In this study, we used partitional clustering methods, as these are better for a large number of variables and computationally faster than hierarchical clustering (if k is small). They may also produce tighter clusters than hierarchical clustering. We selected K-means and fuzzy K-means clustering methods for customer segmentation.

For a brief description of each method, let us assume that we are interested in clustering N samples with respect to P variables into K clusters. For sample i, $x_i = (x_{i1}, x_{i2}, ..., x_{ip}, ..., x_{iP})$ represents a vector of P characteristic variables. Typically K is unknown, but in this study, as TPO decided to segment their international market into four groups, we set k = 4.

### 2.2.1 K-means clustering algorithm

The K-means method is widely used due to its rapid processing ability of large data. K-means clustering proceeds in the following order: Firstly, K number of observations are randomly selected among all N number of observations according to the number of clusters. They become centres of initial clusters. Secondly, for each of the remaining N–K observations, the nearest cluster in terms of the Euclidean distance with respect to $x_i = (x_{i1}, x_{i2}, ..., x_{ip}, ..., x_{iP})$ is found. After each observation is assigned to the nearest cluster, the centre of the cluster is recomputed. Lastly, after the allocation of all observations, the Euclidean distance between each observation and the cluster's centrepoint is calculated, and whether it is allocated to the nearest cluster or not is confirmed.

### 2.2.2 Fuzzy K-means clustering analysis

Fuzzy set theory was introduced in the 1960s as a way of explaining uncertainty in data structure. Fuzzy K-means (also known as fuzzy c-means) clustering has been investigated and compared to the nonfuzzy clustering method.

Hruschka (1986) and Weber (1996) showed in their empirical studies that fuzzy clustering provided more insight than nonfuzzy clustering in terms of market segment information.

Fuzzy clustering segments the samples into $1 < K < N$ clusters, estimates sample cluster membership and simultaneously estimates the cluster centres. The cluster membership of $x_i$ in the cluster $s$; $u_{si}$ is between 0 and 1 and is defined as follows:

$$u_{si} = \frac{1}{\sum_{j=1}^{K} \left( \frac{\|x_i - v_s\|^{2/(m-1)}}{\|x_i - v_j\|^{2/(m-1)}} \right)}, \text{ for } x_i \neq v_j, \forall s, i, \text{ and } m > 1 \tag{4}$$

where $m$ is the smoothing parameter which controls the fuzziness of the clusters, and $v_s$ is the vector of cluster centres $(v_{s1}, v_{s2}, ..., v_{sp}, ..., v_{sP})$, defined as:

$$v_s = \frac{\sum_{I=1}^{N} (u_{si})^m x_i}{\sum_{i=1}^{N} (u_{si})^m} \quad \forall s. \tag{5}$$

The optimal value of $u$ is obtained so as to minimise the following objective function:

$$Min \sum_{i=1}^{N} \sum_{s=1}^{K} (u_{si})^m (\|x_i - v_s\|). \tag{6}$$

The constraints used are as follows:

$$0 \leq u_{si} \leq 1, \qquad \forall s, i \tag{7}$$

$$\sum_{s=1}^{K} u_{si} = 1. \tag{8}$$

Condition (7) ensures that the degrees of memberships are between 0 and 1, and Condition (8) means that, for a given sample, the degrees of membership across the clusters sum to 1. Once the optimal values of *u* are found, the case with the highest associated *u* is assigned a corresponding cluster (Shina and Sohnb, 2004).

## 2.3  *Comparing the performances of the clustering methods*

There are many different criteria for evaluating the quality of clusters. The best clustering is the one wherein the within-cluster homogeneity (or similarity) and between-cluster heterogeneity (or dissimilarity) are the greatest. All clustering quality criteria are based on this concept. Some of them only satisfy one of these two parameters and some consider both for evaluation purposes (Ye, 2003).

In this section, after defining the linkage rules, we set three criteria for comparing the quality of the clustering methods.

### 2.3.1  *Linkage rules*

There are different rules for calculating distance in clustering. The important ones are as follows:

- *Single linkage (nearest neighbour)* – In this method the distance between two clusters is determined by the distance between the two closest objects (nearest neighbours) in the different clusters.

- *Complete linkage (farthest neighbour)* – In this method, the distance between clusters is determined by the greatest distance between any two objects in the different clusters.

- *Pair-group average* – In this method, the distance between two clusters is calculated as the average distance between all pairs of objects in the two clusters.

- *Pair-group centroid* – The centroid of a cluster is the average point in the multidimensional space defined by the dimensions. In a sense, it is the centre of gravity for the particular cluster. In this method, the distance between two clusters is determined as the difference between centroids (Soman *et al*., 2006).

### 2.3.2  *Clustering quality criteria*

In this section three criteria are defined. The first two are taken from the clustering literature and the third one is defined especially for this study.

*Criterion 1:* The objective of the clustering algorithm is to maximise the sum of the similarities between a cluster centre and all countries in the same cluster, and to minimise the sum of the similarities between two cluster centres in different clusters. Let

$O = \{c^n \mid n = 1, ..., K\}$ be the set of K cluster centres, $c^n$ be the cluster centre of the *n*-th cluster, $O^n = \{x_i \mid i = 1, ..., \|T^c - O\|\}$ be the set of the remaining countries that were not selected as a cluster centre and $T^c$ be the aggregated records on which clustering is performed. Therefore, the quality of the clustering result with K clusters can be defined as Equation (9):

$$\rho(K) = \frac{1}{K} \sum_{n=1}^{K} \left( \underset{1 \le m \le K, m \ne n}{Min} \left\{ \frac{\eta_n + \eta_m}{\delta_{nm}} \right\} \right) \tag{9}$$

$$\eta_n = \frac{1}{\|O^n\|} \sum_{x_i \in O^n} Sim(x_i, c^n) \tag{10}$$

$$\eta_m = \frac{1}{\|O^m\|} \sum_{x_j \in O^m} Sim(x_j, c^m) \tag{11}$$

$$\delta_{nm} = Sim(c^n, c^m). \tag{12}$$

Equation (10) defines $\eta_n$ as the average of similarities between cluster centre $c^n$ and all customers in cluster $O^n$ Equation (11) states that $\eta_m$ is the average of the similarities between cluster centre $c^m$ and all customers in cluster $O^m$. Equation (12) defines $\delta_{nm}$ as the similarity between $c^n$ and $c^m$. $\rho(K)$ is calculated considering the Clustering Quality Function, which is presented in Equation (9) between the lower boundary s and the higher boundary t. The K value, which results in the maximum $\rho(K)$ amount, is the optimal number of clusters.

$$\hat{K} \equiv \arg \underset{s \le K \le t}{Max} \{\rho(K)\}. \tag{13}$$

Using Equation (13), an optimal value for K can be objectively determined for country segmentation (Tsai and Chiu, 2004).

*Criterion 2:* In 1997, Michaud presented a method, called 'intraclass' method, for comparing the performances of different clustering techniques. Intraclass inertia is a measure of how compact each cluster is when the number of clusters is fixed. Usually, the variables are scaled to be in the same range (Dibb and Simkin, 1996). Since we have discrete data in this project, we replace the means of the cluster with its centre. In order to calculate the compactness of each cluster, the intraclass inertia $I_n$ of cluster n is defined as:

$$I_n = \sum_{x_i \in O^n_j} Dist(x_i, c^n). \tag{14}$$

Finally, the intraclass inertia F(K) for a given K cluster is defined as:

$$F(K) = \frac{1}{K} \sum_{n=1}^{K} \sum_{x_i \in O^n} Dist(x_i, c^n). \tag{15}$$

One can see that F(K) is the average squared Euclidean distance between each observation and its cluster mean. The lower the F(K), the more compact the clusters are and the better the clustering.

*Criterion 3:* The criteria below are defined to evaluate clusters' homogeneity and between-clusters heterogeneity at the same time:

$$Q(K) = \frac{1}{K}\left(\frac{IntraDist}{ExtraDist}\right). \tag{16}$$

IntraDist is the sum of the maximum complete linkages between the members of each cluster:

$$IntraDist = \sum_{n=1}^{K} Max(Dist(x_i, x_j)), x_i, x_j \in O^n. \tag{17}$$

ExtraDist is the sum of the minimum single linkages between clusters:

$$ExtraDist = \sum_{n=1}^{K} \underset{1 \leq m \leq K, m \neq n}{(Min(Dist(c_i, c_j))), c_i \in O^n, c_j \in O^m}. \tag{18}$$

This criterion calculates the proportion of the maximum distance of the members in each cluster to the minimum distance between clusters.

## 3   Cluster analysis

In this section, after calculating the value of each cluster, the countries are classified through a decision tree and a customer pyramid was drawn. Consequently, the clusters are analysed based on the classification.

### 3.1   Calculating customer value based on the RFM model

Based on the RFM model, the value of a country can be represented as:

$$V(x_i) = W^R \times R(x_i) + W^F \times F(x_i) + W^M \times M(x_i),$$

where $R(x_i)$, $F(x_i)$ and $M(x_i)$ represent the scores for country $x_i$ in terms of the RFM criteria, respectively. $W^R$, $W^F$ and $W^M$ represent the importance weights for the RFM criteria, respectively. In addition, $W^R + W^F + W^M = 1$.

The profitability of the $n$-th country cluster $O^n$ can be acquired by calculating the average for all country values in the cluster. This can be defined as in Equation (19):

$$V(O^n) = W^R \times R(O^n) + W^F \times F(O^n) + W^M \times M(O^n) \tag{19}$$

$$R(O^n) = \frac{\sum_{x_i \in O^n} R(x_i)}{\|O^n\|} \tag{20}$$

$$F(O^n) = \frac{\sum_{x_i \in O^n} F(x_i)}{\|O^n\|} \tag{21}$$

$$M(O^n) = \frac{\sum_{x_i \in O^n} M(x_i)}{\|O^n\|} \qquad (22)$$

where $R(O^n)$, $F(O^n)$ and $M(O^n)$ represent the scores for the *n*-th cluster $O^n$ in terms of R, F and M, respectively. After the profitability for all clusters is known, the clusters are ranked and the most important one is identified. This is helpful for an enterprise in planning and determining long-time and short-time strategies to offer better service to specific customer clusters.

### 3.2   Customer pyramid

The customer pyramid paradigm provides a company with a mechanism for segmenting its customer base and, in so doing, for visualising and analysing customer behaviour, loyalty and value within each of those customer segments. The customer pyramid is stacked according to this rule: 80% of revenue comes from 20% of customers (Figure 2).

**Figure 2**   Customer pyramid
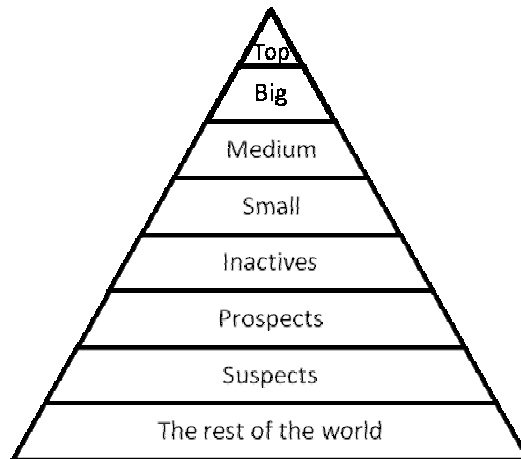


Active customers have transactions with a company and are ranked according to their value in four classes: Top, Big, Medium and Small. 'Value' is defined according to company values.

Inactive customers have accounts with the company but rarely make transactions. Prospects are customers who have had accounts with the company before, but they all are closed now. Suspects are customers who have never had accounts with the company, but it is estimated that that would happen.

The CRM concept can be described in three segments according to the customer pyramid:

1   Get new customers into the pyramid.

2   Move customers higher into the pyramid.

3   Keep the customers in the pyramid (Zeithaml *et al.*, 2001).

## 3.3 Classification

Classification trees constitute a methodology to classify data into discrete units using the tree-structured algorithm. The main purpose of the decision tree is to expose the structural information contained in the data (Soman *et al*., 2006).

In practice, we need threshold values to classify the four different groups. We use decision trees to find the threshold values for customer segmentation.

## 4 Case study

To demonstrate the performance of the proposed country segmentation methodology, we use the export data of the specified 14-year period from the TPO.

According to the retrieved information from TPO databases, the Islamic republic of Iran exports goods and services to a total of 210 countries. These goods and services are wide ranging, about 16 000 types. In this study, the goods and services are categorised into 99 export-group commodities based on the HS code system. There were 222 078 transactions generated jointly by 210 countries in transaction database containing the 99 export-group commodities.

The RFM criteria should be calculated for all the countries and clusters. $Q^R$ is between 1 and 14 and when $Q^R = 14$, it means that Iran has recently exported goods/services to the country concerned. $R(x_i)$ is calculated using Equation (1) and the values are between 0 and 1. $F(x_i)$ is calculated using Equation (2) and, similarly, the values are between 0 and 1. In order to calculate each country's monetary value, we need to calculate the total export volume to that particular country within the specified period. Using Equation (3), $M(x_i)$ is acquired.

### 4.1 Segmentation

Clustering methods (K-means and fuzzy K-means) are used to segment the countries. The RFM variables are used for segmentation. For comparison purposes, the resulting compactness of the clusters by the three mentioned criteria, using Equations (9), (15) and (16), for the two clustering methods (K-means, fuzzy K-means) is summarised in Table 1. In our experiment, both clustering methods are run using Weka software on a Pentium IV, 1.6 Hz PC and the calculation of the criteria is performed using Visual Basic 6.

**Table 1** Compression of each clustering method

| Criteria \ Clustering algorithm | K-means | Fuzzy K-means |
|---|---|---|
| Criterion 1 | 0.33709 | 2.96977 |
| Criterion 2 | 1.03167 | 17.37976 |
| Criterion 3 | 61.25640 | 6850.75301 |

The lower all three criteria are, the better the quality of clustering. As data presents no overlapping and the clusters are disjoint, the clustering by the K-means algorithm has better quality.

## *4.2 Segment analysis*

To provide a clear view for marketing programmes, a three-dimentional RFM model graph is depicted in Figure 3.

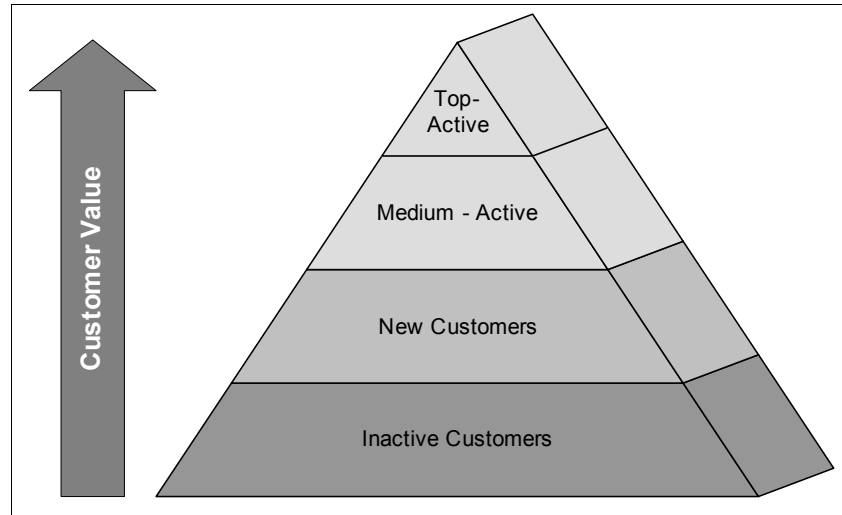**Figure 3**    The three-dimensional RFM model graph (see online version for colours)
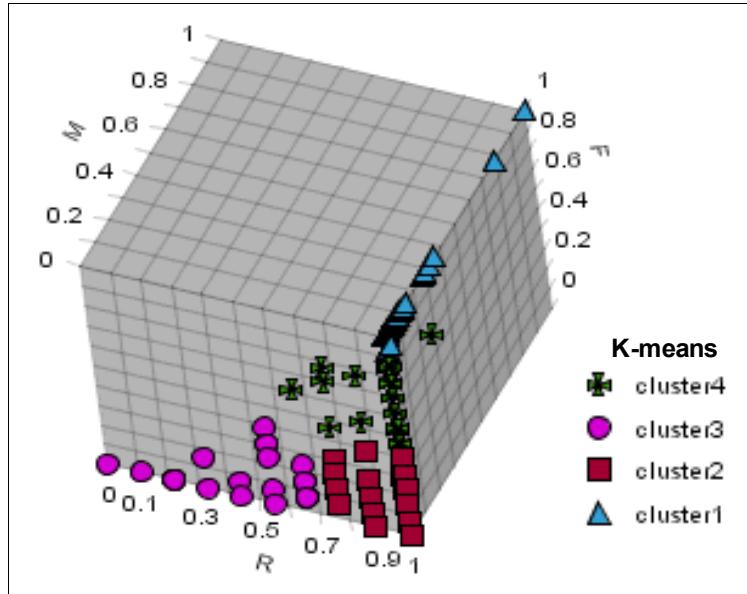


According to experts, the frequency of exports is of higher importance than the recency and monetary value; therefore the F value is greater than R and M in this case study. Consequently, the weights for the RFM criteria were set as $W^R = 0.2$, $W^F = 0.5$ and $W^M = 0.3$.

After a series of calculations using Equations (19) to (22), the value of each cluster is determined and the results are summarised in Table 2.

**Table 2**    Cluster values

| Cluster number | Number of cluster members | $R(O^n)$ | $F(O^n)$ | $M(O^n)$ | $V(O^n)$ |
|---|---|---|---|---|---|
| 1 | 98 | 1.00000 | 0.99184 | 0.06311 | 0.71485 |
| 4 | 49 | 0.97408 | 0.67980 | 0.00635 | 0.53662 |
| 2 | 41 | 0.94878 | 0.16537 | 0.00045 | 0.27258 |
| 3 | 22 | 0.44500 | 0.10182 | 0.00010 | 0.13994 |

We then use the Classification and Regression Tree (C&R) algorithm to demonstrate the four groups. The trees in Figure 4 show the threshold values for customer segmentations and by recalculating R, F and M, we can predict the group of countries and specify their group changes for subsequent years. From Figure 4, if the F value is greater than 0.88, the cluster is defined as Group 1, which means the members are loyal customers. Additionally, according to Table 2 and the RFM model, they are the most valuable customers that Iran has recently had a transaction with. Therefore, Group 1 customers

are defined as Top-active customers at the highest level of the foreign customer pyramid. Also, if the duration of Irans' export is 6 to 12 years, countries are defined as Medium-active customers and named as Group 4. This group aligns in second level of the pyramid. For others, if the R value is greater than 0.715, it means that they are new customers, defined as Group 2. Otherwise they are inactive customers that have recently cut off their relationship with Iran and are known as Group 3. According to this country segmentation based on R, F and M, the Foreign Customer Pyramid is developed (Figure 5).

**Figure 4**    Classifying the countries (see online version for colours)

**Figure 5**    Foreign customer pyramid



We analyse the result of the country segmentation based on R, F and M and determine CRM strategies for each cluster as follows:

Group 1 is the most valuable cluster, whose members have the highest monetary value. Iran has a long-term relationship with these countries, which has continued to the very end of the specified period of analysis. So these are loyal and active customers of ours with the highest profitability level, whom we should keep and retain; we must avoid losing any of them.

The objects of Group 2 are those we have very recent relationships with. The frequency of exports to these countries is very low and we do not have a long-standing relationship with them in our records. Hence, these are the new customers. We have to get to know them more closely in order to grow the volume of exports to them and make the relationship long term and more beneficial.

The members of Group 3 are those we had previous relationships with, but no transaction has been done recently. In other words, these are inactive customers who has churn and are unfortunately great in number. We have to investigate the causes and determine suitable strategies to avoid losing them.

The members of Group 4 are very similar to those of Group 1 in view of recency and frequency, but with a dramatic difference in monetary value, which has made them the second most important cluster in this research. Considering our relationship with these countries, it is quite possible that we can increase the volume of exports by applying appropriate strategies.

## 5    Conclusion

Customer segmentation is critical for a good marketing and CRM programme.

The mass marketing approach cannot satisfy today's diverse customer needs. This diversity should be exploited using market segmentation that divides the market into customer clusters with similar needs, characteristics and purchasing behaviours. This paper introduced a novel country segmentation based on the RFM methodology.

In this article, by using three different criteria, we found that K-means clustering is more stable in segmenting customers than fuzzy K-means and used it to classify four tiers of customers (Top-active, Medium-active, New customer and Inactive) based on the three variables of the RFM model. Then the country segmentation based on RFM was analysed by decision tree, customer pyramid and customer value. Consequently they were used to analyse the relative profitability of each customer cluster and determine the proper strategy for them.

In the future, we may calculate the R, F and M during the monthly duration (but not yearly). Additionally, we can use other clustering algorithms to increase the clustering quality.

## References

Berson, A., Smith, S. and Thearling, K. (2001) *Bulding Data Mining Application for CRM*, McGraw-Hill.

Dibb, S. and Simkin, L. (1996) *The Market Segmentation Workbook: Target Marketing for Marketing Managers*, London: Routledge.

Golsefid, S.M., Ghazanfari, M. and Alizadeh, S. (2007) 'Customer segmentation in foreign trade based on clustering algorithms case study: trade promotion organization of Iran', *International Journal of Computer, Information, and Systems Science, and Engineering*, Vol. 1, No. 3, pp.175–181.

Helsen, K., Jedidi, K. and DeSarbo, W.S. (1993) 'A new approach to country segmentation utilizing multinational diffusion patterns', *Journal of Marketing*, pp.63–71.

Hruschka, H. (1986) 'Market definition and segmentation using fuzzy clustering methods', *International Journal of Research in Marketing*, Vol. 3, pp.117–134.

Hwang, H., Jung, T. and Suh, E. (2004) 'An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry', *Expert Systems with Applications*, Vol. 26, pp.181–188.

Jonkera, J-J., Piersmab, N. and Van den Poelc, D. (2004) 'Joint optimization of customer segmentation and marketing policy to maximize long-term profitability', *Expert Systems with Applications*, Vol. 27, pp.159–168.

Michaud, P. (1997) 'Clustering techniques', *Future Generation Computer System*, Vol. 13, No. 2, pp.135–147.

Shina, H.W. and Sohnb, S.Y. (2004) 'Segmentation of stock trading customers according to potential value', *Expert Systems with Applications*, Vol. 27, pp.27–33.

Soman, K.P., Diwakar, S. and Ajay, V. (2006) *Insight into Data Mining (Theory and Practice)*, India: Prentice-Hall.

Tsai, C-Y. and Chiu, C-C. (2004) 'A purchase-based market segmentation methodology', *Expert Systems with Applications*, Vol. 27, pp.265–276.

Weber, R. (1996) 'Customer segmentation for banks and insurance groups with fuzzy clustering techniques', in J.F. Baldwin (Ed.) *Fuzzy Logic*, New York: Wiley.

Ye, N. (2003) *The Handbook of Data Mining*, Mahwah, NJ/London: Lawrence Erlbaum Associates, Publishers.

Zeithaml, V.A., Rust, R.T. and Lemon, K.N. (2001) 'The customer pyramid: creating and serving profitable customers', *California Management Review*.

## Bibliography

Kamakura, W.A., Wedel, M., de Rosa, F. and Mazzon, J.A. (2003) 'Cross-selling through database marketing: a mixed data factor analyzer for data augmentation and prediction', *Intern J of Research in Marketing*, Vol. 20, pp.45–65.

Kim, S-Y., Jung, T-S., Suh, E-H. and Hwang, H-S. (2006) 'Customer segmentation and strategy development based on customer lifetime value: a case study', *Expert Systems with Applications*, Vol. 31, pp.101–107.

Rygielski, C., Wang, J-C. and Yen, D.C. (2002) 'Data mining techniques for customer relationship management', *Technology in Society*, Vol. 24, pp.483–502.

Wang, H-F. and Hong, W-K. (2005) *Managing Customer Profitability in a Competitive Market by Continuous Data Mining*, Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Hsinchu, Taiwan, ROC.

Wilcox, P.A. and Gurau, C. (2003) 'Business modeling with UML: the implementation of CRM system for online retailing', *Journal of Retailing and Consumer Services*, Vol. 10, pp.181–191.