International Journal of Future Generation Communication and Networking
Vol. 13, No. 4, (2020), pp. 1214–1225

# Presenting a Model for Segmentation, Ranking, and Identification of the Behavioral Patterns of Banking Industry Customers Using Artificial Intelligence and Machine Learning Algorithms

Mohammad paridari[1], Hassan saberi[2*], Zainolabedin amini[3], Ehsan sadeh[4]

1. *PhD student in Business Management, , Department of management, college of human science, saveh Branch, Islamic Azad university, Saveh, Iran*
2. *Assistant Professor in Business Management, Department of management, college of human science, saveh Branch, Islamic Azad university, Saveh, Iran*
3. *Assistant Professor in Public Management, , Department of management, college of human science, saveh Branch, Islamic Azad university, Saveh, Iran*
4. *Assistant Professor in Industrial Management, Department of management, college of human science, saveh Branch, Islamic Azad university, Saveh, Iran*

***Abstract***

*The study used artificial intelligence algorithms for segmentation, ranking, and identification of customer behavior patterns in banking industry and tried to design and present a proper model to identify and rank bank customers given the large volume of customer information in the databases of banks. The study considered the transaction data of Bank Refah Kargaran customers as the population and the transactions performed from September 23, 2018 to March 20, 2019 were selected as the sample. Then preprocessing was done on them and data modeling was done to segment and rank customers using K-means and Self-Organizing Map (SOM) algorithms. Ultimately, the rules governing the relationships between customers' clusters and account types were discovered using Apriori algorithm.*

***Keywords:*** *segmentation, ranking, artificial intelligence, machine learning, K-means, SOM, Apriori*

## 1. Introduction

Nowadays, identifying value and profitability, segmentation, and ranking customers are critical elements for the banks. For identifying and attracting new customers, retaining existing customers, enhancing customer relationship and predicting the future trend of this relationship, the banks use data mining techniques, where feasibility, technical, financial and economic studies should be done. Bhambri (2012) the large number of customers and the specific behavioral characteristics of each of them, the multiplicity and variety of banking products and services, the unattractiveness and inefficiency of mass marketing, and the high costs of implementing personalized marketing as well as external environmental factors affecting banks' performance, such as intense competition, such as the increasing market share of private banks and financial institutions, the rapid growth of technology and e-banking, and so on are the most significant factors in intensifying this challenge.

Moreover, inattention to this issue will cause serious risk, including valuable customers leaving, increase in costs to attract new customers, as well as the need to retain old customers, and so on in banks, preventing which calls for a strong and consistent database and using a management support

* Corresponding Author



ISSN: 2233-7857 IJFGCN

system to optimize customer-related decisions or better customer relationship management (Edelstein, 2003; Smit, 2006; He, Shi, Wan, & Zhao, 2014). Customer segmentation is the most significant approach in retaining customers. By calculating customer lifetime value (CLV), one can determine the value of each customer and where the customer is located (D. Liu & Shih, 2005). The most significant criterion to be used to calculate CLV is the RFM criterion. WRFM model is the optimized RFM model, where W shows the weight for each of the indices (Lemon & Mark, 2006). According to Levitt, if the organization ignores market segmentation, it has basically ignored the market. This means that without a clear understanding of the various requirements and needs of various groups in the market, no organization can be customer-oriented which is essential to market survival (Levitt, 1960).

This study uses data mining techniques to segment, rank, and identify customer behavior patterns for banks given the lack of a proper model for identifying and ranking customers in the banks and present a proper model to identify and rank bank customers.

## 2. A Review of the Theoretical Framework

### 2.1. Segmentation

Market segmentation - which is the process of defining and dividing a large homogeneous market into clear and distinct segments with similar needs and requirements - is a significant part of business strategy. If this is implemented correctly and properly, it supports the overall value of the company with positive effects on sales, profits and market share (Kannisto, 2016). Market segmentation is a strategy that involves dividing a large market into segments where customers have shared requirements and applications for the products and services offered in the market. These parts can be identified through various features according to the composition of each group. One of the major reasons for creating various segments of the market is to learn more about customers and to use various strategies to reach more customer satisfaction and increase profits. The customers in each segment should be similar to each other as much as possible. This allows the company to use various product marketing strategies to meet customer needs and generate more profit (Kieu, Ou, & Cai, 2018); Huerta-Muñoz, Ríos-Mercado, & Ruiz, 2017). Clustering can be defined as a technique where groups have similar features (Y. Liu, Kiang, & Brusco, 2012).

### 2.2. Clustering

Clustering is dividing data into similar groups. The data are clustered according to the principle of maximizing intra-group similarity and minimizing inter-groups similarity. Clustering is a common descriptive approach trying to identify a limited number of clusters to describe the data (KAUFMAN & ROUSSEEUW, 1990). Cluster analysis or clustering is research area defined by a group of objects or targets called clusters or segments, so that the members of one cluster are similar to each other and different from other clusters. Clustering techniques have increased in many study fields like machine learning, statistics, bioinformatics and marketing (e.g., market segmentation) (Trindade, Dias, & Ambrósio, 2017). Clustering is extensively used in many fields. In the commercial field, one can use it to analyze customer behavior and present an outstanding base for developing a marketing strategy. Moreover, in the field of e-commerce, one can use it to analyze the characteristics of similar customers to present better service to the customer (Zhang, Zhang, & Zhang, 2018)

K-means algorithm is the commonest simple clustering method used for large numbers of large-scale numerical data, which presents an efficient way to classify similar data into a similar cluster (Yu, Chu, Wang, Chan, & Chang, 2018). Kohonen introduced SOM for the first time. These networks operate according to competitive learning and have unobserved training, meaning that they do not need to have expected outputs during the training process (Chuang, Wei, Zhifu, & Zhi, 2017). The input of these networks is a set of training data usually applied to the network for clustering or categorization (Lopez, Valero, Senabre, Aparicio, & Gabaldon, 2012).



ISSN: 2233-7857 IJFGCN

### 2.3. Apriori Algorithm

The purpose of data mining is finding dependency, robust and significant rules. Dependency or association rules were first introduced in 1993 for values of zero and one by Agraval et al. and published in 2002 (Aggarwal, Procopiuc, & Yu, 2002). Apriori algorithm general procedure is as follows:

Any non-empty subset of an iterative set will certainly be iterative. Two major steps shape the general procedure of this algorithm. In one step, the K member set of $C_k$ candidates is created from $L_{k-1}$ bond with itself. Then, in the next step, the number of iterations of each $C_k$ set is calculated and then based on the minimum acceptable iteration criterion, the sets with a small number of iterations are deleted with the sets remaining at this stage forming $L_k$. The two mentioned steps will continue until no new $L_k$ can be found. Hence, the algorithm condition would be empty $L_k$ (Wen, Liao, Chang, & Hsu, 2012).

### 2.4. RFM Model to Calculate CLV

RFM approach, which is suggested by Hughes (Hughes, 2012) is an effective approach of market segmentation in marketing where the audiences are classified according to the analysis of the duration of a period from the last purchase (R), the number of purchases within a given time (F), the amount of money spent during this specified time (M), or recency, frequency, and monetary value (Qin, Yuan, & Wang, 2017). This method is one of the most significant methods for measuring CLV because by segmenting customers using transactional data and based on this model, the valuable customers are singled out (D. Liu & Shih, 2005). Stone argues that each of the criteria - recency, frequency, and monetary value - should be weighed according to the industry type. In his own study, he gave priority to frequency, recency, and finally the monetary value due to the industry type, but he set the priorities intuitively (Stone, 2007). Recency is the time past since the last purchase of the customer. The scholars recognized that the shorter this period, the greater the likelihood of customer response. Frequency is the number of times a customer buys from the beginning to the end of a given period that increases the probability of a response if it is higher. The monetary value is the amount of money paid by customers in the past, or a specific period, or the average purchase of each, where the higher the order, the greater the likelihood of a response. However, the negative and positive relationships stated do not always apply. Nonetheless, when the relationship between recency and response probability, which is usually negative, is determined, the response rate can be calculated using previous customer data. The relationship between frequency and probability of response should also be empirically determined, yet it is usually positive. Frequency is expressed as the number of times a customer buys from the beginning or the number of purchases from the beginning until now (Bult & Wansbeek, 1995)

## 3. Methodology

As the study used transactional and demographic data of the customers simultaneously, it was descriptive-exploratory. In terms of time, the study was retrospective or post hoc, where data collection is done at a specified time at the end of which the data is analyzed. In the study, the transaction data of the customers of Bank Refah Kargaran was used to test the proposed model and the approaches introduced for segmentation and classification. Thus, the transaction data of this bank's customers were considered as the population and the transactions performed from September 23, 2018 to March 20, 2019 were selected as the sample. The data had two main parts. The first part was about demographic information of the customers like gender, age, and so on and the second part, the transaction data of the bank's customers, used as a tool for data collection. Hence, 1123735 transactions were collected and recorded. This data was used to examine the classification and ranking models of customers. Because of using data mining models as well as changing economic conditions and customer preferences, the need was felt to use the most recent data; thus, all of the data for the stated period was used that was relatively complete. According to the points stated, the three variables - recency, frequency, and monetary value are used for market



ISSN: 2233-7857 IJFGCN

segmentation. The date of the last transaction was as the recency; the number of iterations of transactions per customer as the frequency, and the balance of each customer account as monetary value or volume of exchange to get these three variables. The results were analyzed using age, gender, and occupation data. Moreover, for the association rules, the type of customer accounts and clusters obtained were used to discover relationships. Silhouette criterion was used to determine the validity and reliability of the results of segmentation methods and since no questionnaires were used to collect data at this stage, no reliability was needed. The data was extracted directly from precisely stored customer databases. Then, pre-process was done on them to detect possible errors in the data. IBM SPSS Modeler 18 was used to perform data mining calculations in this study.

## 4. Research Findings

### 4.1. Pre-processing Data

The data provided by Bank Refah Kargaran was related to occupation, age, gender, last account turnover, average account balance, debtor turnover, creditor turnover, account description, and account type of customers with transactions during the six months. Thus, all of these features had to be examined separately and their outliers identified, to obtain complete and impeccable data, whose steps are as follows:

- The age range for this study was set at 18 years to 70 years and the rest of the data were excluded from the analysis.
- Code 1 was used to show the males and 2 to show the females.
- As the data transaction date was from September 23, 2018 to March 20, 2019, the data whose date of the last transaction was not in this range was deleted.
- The account and the type of account whose data was incompatible with or otherwise too small were considered outliers and deleted.
- As the frequency of transactions was too large for data analysis, some groupings were done. In this study, groupings were done according to gender, age, and occupation. For instance, 35-year-old men with the occupation code 700 were grouped into one group. Then recency, frequency, and mean transactions were considered as recency, frequency, and exchange volume of that group.

### 4.2. Pairwise Comparison of Analytic Hierarchy Process (AHP), expert Questionnaire Design and Clustering

A pairwise comparison questionnaire (not hourly) of AHP was used by the banking industry experts to determine the weight of the criteria and to rank the options.

The value of each RFM model index was determined by multiplying the normalized index value by its weight. The value of these indices is as follows:

M = WM

F = WF

R = WR

Table 1 indicates that in ranking bank customers, the volume of transactions is 6 times more important than the recency and 3 times more important than the frequency.

**Table 1. Pairwise Comparison and Determining Normalized Weight of Criteria**

| Priority | Transaction | Frequency | Recency | |
|---|---|---|---|---|
| 0.168333 | 0.289 | 0.105 | 0.111 | Recency |
| 0.213667 | 0.19 | 0.112 | 0.333 | Frequency |
| 0.618 | 0.515 | 0.783 | 0.556 | Transaction |

The sum of the average values of RFM indices in that cluster is used to calculate the value of each customer in each cluster:

CLV = WR + WF + WM



ISSN: 2233-7857 IJFGCN

### 4.3. Customer Grouping Using RFM Approach

The rating of the RFM approach is by subtracting the maximum and minimum values of each index and dividing it into ten groups. For instance, for the purchase frequency, the first group includes 10% of the customers with the most purchases in the target period with 9 points assigned, the second 10% receives eight points and this process goes on until zero is assigned to the last 10% with the least frequency. Such a process is adopted for purchasing price index too and for recency, the higher the value, the higher its negative effect will be given its nature. It is because it shows the days when the customer did not have a transaction. Thus, scoring recency index is reversed: zero is assigned to that 10% customer with the highest delay that is the number of days that have not been transacted. This process goes on until nine is assigned to 10% of the customers with the lowest amount of delay. Table 2 shows the final scoring.

**Table 2. Scaling RFM Features**

| Monetary value (M) % | Frequency (F)% | Recency (R)% | |
|---|---|---|---|
| 90-100 | 90-100 | 0-10 | Score 9 |
| 80-90 | 80-90 | 10-20 | Score 8 |
| 70-80 | 70-80 | 20-30 | Score 7 |
| 60-70 | 60-70 | 30-40 | Score 6 |
| 50-60 | 50-60 | 40-50 | Score 5 |
| 40-50 | 40-50 | 50-60 | Score 4 |
| 30-40 | 30-40 | 60-70 | Score 3 |
| 20-30 | 20-30 | 70-80 | Score 2 |
| 10-20 | 10-20 | 80-90 | Score 1 |
| 0-10 | 0-10 | 90-100 | Score 0 |

After calculating and assigning the scores from zero to nine points, the weights obtained are multiplied by each, and by adding up the final scores of three indices for all customers, the WRFM score is obtained for each customer. The higher the WRFM score for one customer, the better it will be.

### 4.4. Modeling customer ranking and CLV

#### 4.4.1. Clustering data using K-MEANS algorithm

One of the significant points of clustering is determining the optimal number of clusters that in most algorithms like K-means, have to be determined by the user himself and no specific way is specified. One possible solution is to test various ks and to determine the optimal value based on a set of predefined indices. Hence, clustering was done for 1 to 10 clusters with K-Means algorithm based on RFM method, and then the silhouette criterion was used to determine the optimal number of clusters. According to the Silhouette criterion, three clusters were selected as the optimal cluster number. The following table shows the value of these variables for the number of clusters 3.

**Table 3. Characteristics of Each Cluster Using k-means Algorithm**

| | **Mean WR** | **Mean WF** | **Mean WM** | **Mean CLV** | **Records** |
|---|---|---|---|---|---|
| **Cluster 1** | 0.83 | 1.27 | 4.29 | 6.39 | 813 |
| **Cluster 2** | 0.5 | 1.52 | 1.81 | 3.83 | 1062 |
| **Cluster 3** | 0.31 | 0.15 | 0.31 | 0.77 | 1203 |

The table shows the mean values of recency, frequency, and monetary value of each section separately, and means CLV of each cluster and the number of customers in each cluster. Based on the information given, the first part with 813 customers has the highest mean CLV and is the most valuable part.



ISSN: 2233-7857 IJFGCN

#### 4.4.2. Data clustering using SOM algorithm

Another approach to examine the optimal number of clusters is to check Kohnen's SOM output or SOM algorithm. Customer rating data (CLV) was given to the network as the input and Kohnen's model output map aspects were considered as 7 × 10 two-dimensional map, which is the initial proposed mode. The color difference in the map shows the number of customers in each home: the more customers in a home, the bolder that home is. By examining the results of this section, one can see that the customers can be divided into three sections.

**Table 4. Characteristics of the Clusters Obtained from SOM Algorithm**

| | Mean WR | Mean WF | Mean WM | Mean CLV | Records |
|---|---|---|---|---|---|
| **Cluster 1** | 0.89 | 1.33 | 4.5 | 6.78 | 724 |
| **Cluster 2** | 0.64 | 1.54 | 1.83 | 4.01 | 1104 |
| **Cluster 3** | 0.17 | 0.12 | 0.24 | 0.53 | 1250 |

In this table, it is seen that the first cluster with 724 customers and mean CLV score of 6.78 have the highest values of recency, frequency, and monetary value.

The values obtained from clustering using SOM algorithm are similar to k-means algorithm. However, as the silhouette criterion for SOM algorithm is lower than that of k-means algorithm, the quality of segmentation of Kohnen's model or SOM is less compared to k-means model.

Recency, frequency, and monetary values of the three parts show the differences between the parts. Customers in the first part have the highest levels of recency, frequency, and monetary values than the other two. These customers can be called high value customers. In contrast, the third part has the lowest values of recency, frequency, and monetary value compared to the other two. Hence, the name of this group of customers is low value customers. The second part with the values of recency, frequency, and monetary value is called the moderate value part.

#### 4.4.3. Cluster Analysis According to RFM

Marcus (1998) presented the customer value matrix, where the customer purchase frequency (F) and customer purchase monetary value (M) form the two vectors of this matrix. In addition, the collaboration time, latency, and recency of purchase are related to customer loyalty. Marcus argues that the longer the duration of the collaboration and the shorter time passed from the last customer transaction, the more loyal the customer will be. In this study, given the lack of customer collaboration time data, this factor was deleted and segmentation was done regardless of the collaboration duration.

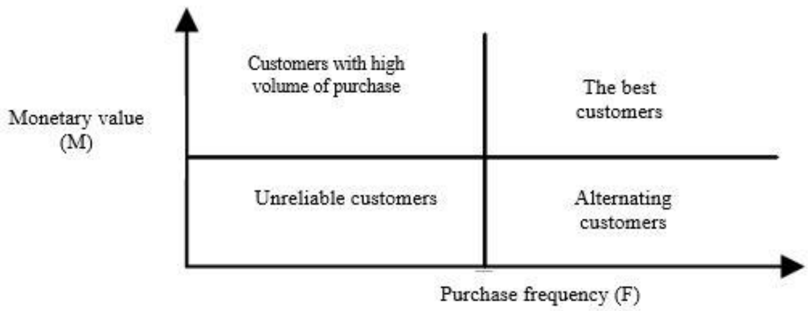


**Figure 2. Customer Value Matrix (Marcus, 1998)**

In this study, according to the customer segmentation description (Ha & Park, 1998), ↑ is used to show a higher group mean compared to the overall mean and ↓ to show the group mean value lower than the overall mean.



ISSN: 2233-7857 IJFGCN

**Table 5: Total mean values of variables R, F and M**

| WR-average | WF-average | WM-average |
|---|---|---|
| 0.51 | 0.92 | 1.88 |

Table 6 is the comparison of the mean values of WR, WF, and WM of the clusters with the total mean of these variables.

**Table 6. Results of the Mean Values of WR, WF, and WM of Each Cluster with Total Mean**

| | Mean WR | Mean WF | Mean WM | Records | Cluster percent |
|---|---|---|---|---|---|
| **Cluster 1** | 0.83↑ | 1.27↑ | 4.29↑ | 813 | 26.4 |
| **Cluster 2** | 0.52↑ | 1.52↑ | 1.81↓ | 1062 | 34.5 |
| **Cluster 3** | 0.3↓ | 0.15↓ | 0.31↓ | 1203 | 39.1 |

The length of bank customer relationship is not available given the incomplete data. However, if we consider that the customer age is directly related to customer relationship length with bank (as the results of previous parts showed, the older the people get, they are placed in better clusters) we can set the age instead of the length of the relationship with the customer. Then, without weighing it, we can examine the customer in terms of the loyalty matrix.

According to Chang and Tsay (2004), by adding L index (customer relationship length), the context is provided for a more detailed analysis of the customers. Moreover, it claims that the longer the customer relationship, the higher the loyalty, and the shorter the latency (latency or recency), the higher the customer loyalty will be. The two indices of customer relationship length (L) and recent transaction time (R) are defined as customer loyalty matrices (Chang & Tsay, 2004)

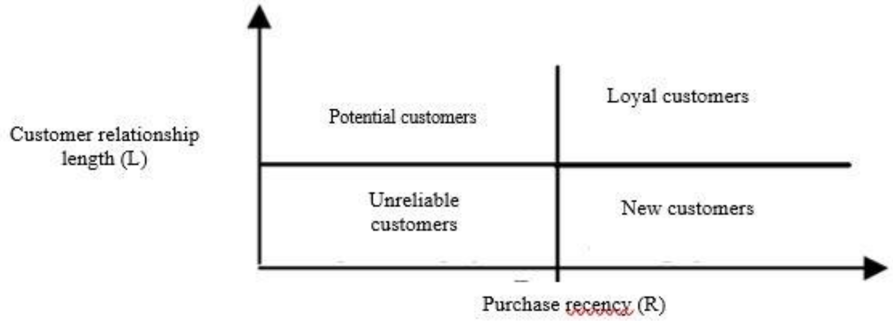


**Figure 3. Customer Loyalty Matrix (Chang & Tsay, 2004)**

If age is considered as L, the age range from 18 to 70 is divided into ten, and each interval is assigned a score of 0 to 9, so the mean will be according to Table 7.

**Table 7. The Mean Values of the Variables L, WR, WF, and WM**

| L-average | WR-average | WF-average | WM-average |
|---|---|---|---|
| 3.5 | 0.51 | 0.92 | 1.88 |



ISSN: 2233-7857 IJFGCN

Table 8 is the comparison of the mean values of L, WR, WF, and WM of the clusters with the total mean of these variables.

**Table 8. The Results of Mean Values of L, WR, WF, and WM of Each Cluster with the Total Mean**

| | Mean L | Mean WR | Mean WF | Mean WM | Records | Cluster percent |
|---|---|---|---|---|---|---|
| **Cluster 1** | 5.5↑ | 0.83↑ | 1.27↑ | 4.29↑ | 813 | 26.4 |
| **Cluster 2** | 3↓ | 0.52↑ | 1.52↑ | 1.81↓ | 1062 | 34.5 |
| **Cluster 3** | 3.2↓ | 0.3↓ | 0.15↓ | 0.31↓ | 1203 | 39.1 |

According to Table 8 in the customer loyalty matrix, the customers in the third cluster are unreliable customers, second cluster customers are new customers, and first cluster customers are loyal ones.

### 4.5. Association Rules (Apriori Algorithm)

As the data obtained from Bank Refah Kargaran includes customer account data as well, the rules governing them should also be explored to determine which accounts have the most transactions and what are the rules governing them. In the data used in the study, 108 types of accounts were used to express customer accounts.

In this section of the study, the customers of each cluster were examined for the accounts type they use, and the accounts that stood for more than one percent of transactions over a six-month period were analyzed using association rules and Apriori algorithm. The following are the rules with minimum support of 0.4% and minimum reliability of 75%.

| Consequent | Antecedent | Support % | Confidence % |
|---|---|---|---|
| Cluster 2 | Account 97<br>Account 25<br>Account 79 | 0.422 | 92.308 |
| Cluster1 | Account 98<br>Account 79 | 0.487 | 86.667 |
| Cluster1 | Account 108<br>Account 51<br>Account 79 | 0.552 | 82.353 |
| Cluster1 | Account 108<br>Account 41 | 0.747 | 78.261 |
| Cluster3 | Account 64 | 28.428 | 75.543 |

**Figure 4. The Rules from Association Rules**

## 5. Research Results

The analyses of the expert questionnaire filled in by the managers and experts of the banking industry indicate that transaction volume is more significant and valuable compared to recency and frequency. Moreover, the results indicate that according to the Silhouette criterion, the optimal number of three clusters is more suitable for customer clustering and the results are as follows:

### 5.1. First Cluster

This cluster has 813 records standing for 26.4% of total customers and is higher than the overall mean in all three indices and the customers are named as the best customers. Moreover, the



ISSN: 2233-7857 IJFGCN

customers in this cluster are the loyal ones, with 53% men and 47% women. The age range of the cluster is 18 to 70 years with the mean age of the cluster as 52 years. Furthermore, 74.5% of the customers are over 45 years old and only 25.5% are in the age group of 18-44 and occupations with code 706, 3, and 70 have the highest percentage in this cluster, respectively. This shows non-medical occupations and business, and the data for code 70 is incomplete and the occupation related to code 70 is not stated in the data obtained from the bank.

### 5.2. Second Cluster

This cluster, with 1062 records, is almost moderate in terms of transaction recency and higher than overall mean in terms of transaction frequency. The volume of transactions that was the most important factor in the weighting of the experts in this cluster was less than the overall mean, and is intermittent among the customers in terms of Marcus customer value matrix. Moreover, the customers in this cluster are the new ones, with 48% of them being men and 52% women. The age range of this cluster is 18 to 70 years and the average age of this cluster is 39 years. Additionally, 38% of these clusters were those under 30 and 30% are 30-45 years of age with the highest percentage and those aged 50-75 years of age have the lowest percentage of this cluster (10%). In this cluster, most of the customers have jobs with code 13, 12, and 2, where code 2 shows self-employed jobs, and codes 12 and 13 are unspecified job titles.

### 5.3. Third Cluster:

This cluster, with 1203 records, has an unsatisfactory state for all three indices of R, F, and M and is lower than the overall mean in all three indices and the customers are considered unreliable in terms of Marcus customer value matrix. Moreover, the customers are unreliable ones, with 48% of them as men and 52% women. The age range of this cluster is 18 to 70 years and the mean age of the population is 42 years. Moreover, 36% of these clusters were those under 30 and 23% were 30-45 years with the highest percentage and those aged 57-70 had the lowest percentage (11.5%). Most of these customers in this cluster have the codes 13, 12, and 2, which code 2 shows self-employed jobs, and codes 12 and 13 are as unspecified job titles in the obtained data.

The first cluster uses more the long-term investment deposit accounts and real welfare plan account type (98), savings account type and account type, customers (79), stage 16 deposit-certificate (108), individuals loan current type and type accounts use current check-in facilities without checklists (51), current debit deposits, and the type of checking account with an electronic card (41). Moreover, the majority of the individuals in this cluster have two or three accounts each. The second cluster of long-term investment deposit account type and account type, new year deposit (97), savings and loan account type, customers (79), and short-term entities and deposit account, short-term deposit (25) and most customers have these three accounts together. The third cluster uses just individuals' current accounts deposits and account types, and help-seekers of the Welfare Organizations (64).

**Acknowledgments**
In the end, thank you to all the friends who helped us write this article

**References**

[1] Aggarwal, C. C., Procopiuc, C., & Yu, P. S. (2002). Finding localized associations in market basket data. *IEEE Transactions on Knowledge and Data Engineering*, *14*(1), 51–62. https://doi.org/10.1109/69.979972

[2] Bhambri, V. (2012). *Volume 2 , Issue 9 ( September 2012 ) IMPLEMENTATION OF DATA MINING IN BANKING SECTOR- A IJRIM*. *2*(9), 21–30.

[3] Bult, J. R., & Wansbeek, T. (1995). Optimal Selection for Direct Mail. *Marketing Science*, *14*(4), 378–394. https://doi.org/10.1287/mksc.14.4.378

[4] Chang, H. ., & Tsay, S. . (2004). Integrating of SOM and K mean in data mining clustering: An empirical study of CRM and profitability evaluation. *Journal of Information Management*,



ISSN: 2233-7857 IJFGCN

*11*(4), 161–203.

[5] Chuang, C., Wei, Z., Zhifu, W., & Zhi, L. (2017). The diagnosis method of stator winding faults in PMSMs based on SOM neural networks. *Energy Procedia*, *105*, 2295–2301. https://doi.org/10.1016/j.egypro.2017.03.663

[6] Edelstein, H. (2003). *Building profitable customer relationships with data mining*.

[7] Ha, S. H., & Park, S. C. (1998). Application of data mining tools to hotel data mart on the Intranet for database marketing. *Expert Systems with Applications*, *15*(1), 1–31. https://doi.org/10.1016/S0957-4174(98)00008-6

[8] Huerta-Muñoz, D. L., Ríos-Mercado, R. Z., & Ruiz, R. (2017). An iterated greedy heuristic for a market segmentation problem with multiple attributes. *European Journal of Operational Research*, *261*(1), 75–87. https://doi.org/10.1016/j.ejor.2017.02.013

[9] Hughes, A. M. (2012). *Strategic Database Marketing 4e : The Masterplan for Starting and Managing a Profitable , Customer-Based Marketing Program*.

[10] Kannisto, P. (2016). “I’M NOT A TARGET MARKET”: Power asymmetries in market segmentation. *Tourism Management Perspectives, 20*, 174–180. https://doi.org/10.1016/j.tmp.2016.09.001

[11] KAUFMAN, L., & ROUSSEEUW, P. J. (1990). *Finding Groups in Data An Introduction to Cluster Analysis*. A JOHN WILEY & SONS.

[12] Kieu, L. M., Ou, Y., & Cai, C. (2018). Large-scale transit market segmentation with spatial-behavioural features. *Transportation Research Part C: Emerging Technologies*, *90*(August 2017), 97–113. https://doi.org/10.1016/j.trc.2018.03.003

[13] Lemon, K. N., & Mark, T. (2006). Customer Lifetime Value as the Basis of Customer Segmentation Customer Lifetime Value as the Basis of Customer Segmentation : Issues and Challenges. *Journal of Relationship Marketing*, (December 2014), 37–41. https://doi.org/10.1300/J366v05n02

[14] Levitt, T. (1960). *Marketing Myopia*. Harvard Business Review. July-August.

[15] Liu, D., & Shih, Y. (2005). *Integrating AHP and data mining for product recommendation based on customer lifetime value*. *42*, 387–400. https://doi.org/10.1016/j.im.2004.01.008

[16] Liu, Y., Kiang, M., & Brusco, M. (2012). A unified framework for market segmentation and its applications. *Expert Systems with Applications*, *39*(11), 10292–10302. https://doi.org/10.1016/j.eswa.2012.02.161

[17] López, M., Valero, S., Senabre, C., Aparicio, J., & Gabaldon, A. (2012). Application of SOM neural networks to short-term load forecasting: The Spanish electricity market case study. *Electric Power Systems Research*, *91*, 18–27. https://doi.org/10.1016/j.epsr.2012.04.009

[18] Marcus, C. (1998). A practical yet meaningful approach to customer segmentation. *Journal of Consumer Marketing*, *15*(5), 494–504.

[19] Qin, R., Yuan, Y., & Wang, F. (2017). Exploring the Optimal Granularity for Market Segmentation in RTB Advertising via Computational Experiment Approach. *Electronic Commerce Research and Applications*. https://doi.org/10.1016/j.elerap.2017.07.001

[20] Stone, B. (2007). Successful Direct Marketing Methods. In *Journal of Advertising* (Eighth Edi). Lincolnwood: NTC Business Books, IL.

[21] Trindade, G., Dias, J. G., & Ambrósio, J. (2017). Extracting clusters from aggregate panel data: A market segmentation study. *Applied Mathematics and Computation*, *296*, 277–288. https://doi.org/10.1016/j.amc.2016.10.012

[22] Wen, C. H., Liao, S. H., Chang, W. L., & Hsu, P. Y. (2012). Mining shopping behavior in the Taiwan luxury products market. *Expert Systems with Applications*, *39*(12), 11257–11268. https://doi.org/10.1016/j.eswa.2012.03.072

[23] Yu, S. S., Chu, S. W., Wang, C. M., Chan, Y. K., & Chang, T. C. (2018). Two improved k-means algorithms. *Applied Soft Computing Journal, 68*, 747–755. https://doi.org/10.1016/j.asoc.2017.08.032

[24] Zhang, G., Zhang, C., & Zhang, H. (2018). Improved K-means Algorithm Based on Density Canopy. *Knowledge-Based Systems*. https://doi.org/10.1016/j.knosys.2018.01.031



ISSN: 2233-7857 IJFGCN

International Journal of Future Generation Communication and Networking
Vol. 13, No. 4, (2020), pp. 1214–1225

## Authors

Author's picture should be in grayscale.

Picture size should be absolute 3.18cm in height and absolute 2.65cm in width

**Mohammad paridari, Hassan saberi, Zainolabedin amini, Ehsan sadeh,** Author's profile.

**Mohammad paridari**

**Hassan saberi**

**Zainolabedin amini**



ISSN: 2233-7857 IJFGCN

**Ehsan sadeh**



ISSN: 2233-7857 IJFGCN