International Business Management 10 (8): 1406-1413, 2016
ISSN: 1993-5250

Clustering Customers by C-Mean Method (Case Study: Golestan Company)

[1]Mansour Momeni, [1]Sahar Yazdani and [2]Maryam Faridi Khorshidi
[1]University of Tehran, Tehran, Iran
[2]University of Tehran (Alborz Campus), Tehran, Iran

Abstract: Segmentation in the form of a managerial philosophy is based on micro-economy theory and tendency to customer requirements and is one of the most fundamental concepts in new marketing. Market segmentation includes the view to a heterogeneous market as some smaller homogenous markets in response to various preferences of product among important sectors of market. Today, companies are obliged to recognize the customers and planning for them. This study applied fuzzy clustering and separated customers of Golestan Company. The study was performed on 1589 samples of customers of Golestan Company in Kerman Province and two clusters of customers were identified by Silhouette coefficient. Then, membership degree of each object was achieved for total customers and separated cash and non-cash customers by Matlab Software. Finally, the statistical feature of each cluster was determined by SPSS Software. The result of this study helped the analysis of behavior of customers and selection of suitable policies of each cluster.

Key words: Clustering, c-mean clustering, data mining, Silhouette index, philosophy

INTRODUCTION

Today, the role of customers is turned from following the manufacturer to guiding the investors, producers and even researchers and innovators. Thus, organizations required identification of their customers and planning for them. Based on tense competition in market and various choices of products and services of customers, correct recognition of their behavior is the most important dimensions of customer relationship management (Berson *et al*., 2000). In recent years, considering customer satisfaction and fulfilling his needs are increased (Hsu *et al*., 2006). Indeed, customer satisfaction is an introduction for customer loyalty retention (Jamal and Anastasiadou, 2009). Thus, the analysis of their behavior and selection of suitable marketing method based on these analyses is an important factor for enterprises survival.

The organizations should be able to understand which group of customers are the best and which goods and services fulfill the needs of selected group. Some of statistical methods and machine learning are used for this purpose and these methods have some limitations alone. Data mining is the tool of processing considerable data in data basis and it can convert data to useful information. Using customers' segmentation as one of data mining techniques is useful to divide heterogeneous customers into homogenous groups to recognize the customer behavior. Formulation of suitable marketing strategy and effective communication with each section of customers is useful to attract the satisfaction and retention of customers, absorption of new customers and survival of these organizations.

Golestan Company is a production and packing company of food products in Iran with investment of Golestan Business Group (GBG). This company was established in 1953 as the first tea package and distribution in Iran, thus, the founder of Golestan Company is the father of marketing and tea distribution in Iran. After half a century efforts, Golestan Company attempts to apply its experiences and produce reliable and high quality goods. The activity of this company is in channel distribution of food by equipped distribution system and update technology and customers' orders are distributed. Unique warehousing of distribution organization in goods maintenance is provided in central warehouse and tens of regional warehouses all over Iran and goods distribution is provided in >100000 users all over Iran. Data basis of factories and food manufacturing companies in Kerman Province includes pasta, reshtefarangi, reshte ash, wafer, biscuite, sause, tomato paste, sugar, juice, cake, oil, flower oil, herbal distilled, tea, bread, potato chips, syrup, jam, cookies, etc. Golestan Company attempts to observe global standards

Corresponding Author: Sahar Yazdani, University of Tehran, Tehran, Iran

in production and present appropriate Iranian product services to market. This company acquired various certificates including ISO 9001 and HACCP from TUV of germany.

This study attempted to use clustering method and presented a framework to identify customers of Golestan Company. The results of study helped Golestan Company to clustering the customers and take different policies and strategies to each of customers based on the cluster belonging to the customer. They evaluated the needs of their customers based on their clustering and responded separately and this led into maintaining satisfaction and loyalty of customers in company. This study aimed to identify customers, distinction between them, determining valuable customers, absorbing valuable customers and their retention. To create a ranking system we should use segmentation techniques in data mining. Thus, we divided the customers of this company into some groups based on four criteria of number of orders of each customer in 2013 of company, annual purchase price of each customer in 2013, payment method of purchase of each customer in 2013 and number of bounced checks of each customer in 2013. After the identification of groups they are clustered and clustering is done by c-mean method.

Literature review: In recent years, the electronic data reservoir of commercial institutes and companies is increased as many of these resources has made impossible their analysis for human being to create information or useful models for decision making process. Some techniques as data mining can automatically extract information among the great amount of data (Dastgir and Sardasht, 2011). Data mining is the exploration and analysis process by automatic and semi-automatic tools of great amount of data to detect significant and basic models (Rygielski *et al*., 2002). In other words, information and knowledge extraction and detection of hidden models is defined from a great and complex data basis (Berson *et al*., 2000). Figure 1 defines different types of predictor and descriptive data mining methods.

Data mining is applied as computation tool for clustering stakeholders and actors in policy making processes and decision making. Clustering analysis is a sample of undirected data mining methods organizing a set of data in a series of clusters as the data (as showing vector of quantitative values in a multi-dimensional space) in each cluster, it has the highest similarity degree and data of various clusters have maximum non-similarity degree. In each clustering process, at first data matrix is provided and is standardized by special methods. Table 1 shows some of standardization equations.

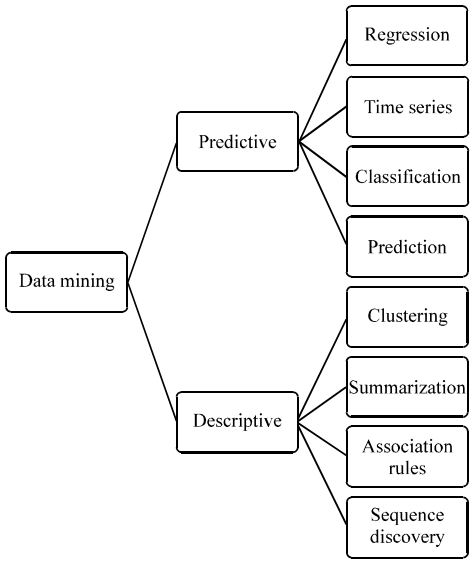

Fig. 1: Different types of activies and application of data minig

Table 1: Methods of data matrix standardization

Function	Equations
Standard normal	$Z_{ij} = \frac{X_{ij} - \bar{X}_i}{S_i}$
Second function	$Z_{ij} = \frac{X_{ij}}{S_i}$
First rank function	$Z_{ij} = \frac{X_{ij}}{RMAX_i}$
Second rank function	$Z_{ij} = \frac{X_{ij} - RMIM_i}{RMAX_i - RMIN_i}$
Normalizing function	$Z_{ij} = \frac{X_{ij}}{\sum_{j=1}^{t} X_{ij}}$
Rank function	$Z_{ij} = RANK\ (X_{ij})$

Table 2: Some of distance criteria for quantitative indices

Distance function	Formula and comments
Euclidean distance	$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$
Hamming (city block) distance	$d(x, y) = \sum_{i=1}^{n} \lvert x_i - y_i \rvert$
Tchebyschev distance	$d(x, y) = max_{i=1,2,..,n} \lvert x_i - y_i \rvert$
Minkowski distance	$d(x, y) = \sqrt[p]{\sum_{i=1}^{n} (x_i - y_i)^p}$, p>0
Canberra distance	$d(x, y) = \sum_{i=1}^{n} \frac{\lvert x_i - y_i \rvert}{x_i + y_i}$, xi and yi are positive
Angular separation	$d(x, y) = \frac{\sum_{i=1}^{n} x_i y_i}{\left[\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^2\right]^{1/2}}$

After standardization, proximity matrix (distance or similarity) is computed. In Table 2 and 3 some examples of distance and similarity criteria for quantitative indices are shown. One of clustering methods is implemented. In final stage, clustering validity is evaluated.

Table 3: Some similarity/dissimilarity coefficients for quantitative indices

Name	Equations
Cosine coefficient	$\frac{\sum_{i=1}^{d} x_i y_i}{\sqrt{\sum_{i=1}^{d} x_j^2}\sqrt{\sum_{i=1}^{d} y_j^2}}$
Pearson correlation coefficient	$\frac{\sum_{i=1}^{d} x_i y_i - n\bar{x}\,\bar{y}}{\sqrt{\sum_{i=1}^{d} x_j^2 - n\bar{x}^2}\sqrt{\sum_{i=1}^{d} y_j^2 - n\bar{y}^2}}$
Canberra coefficient	$\frac{1}{d}\sum_{i=1}^{d}\frac{\lvert x_j - y_j\rvert}{(x_j + y_j)}$
Bray-Curtis coefficient	$\frac{\sum_{j=1}^{d}\lvert x_j - y_j\rvert}{\sum_{j=1}^{d}(x_j + y_j)}$
Johnson-wichern coefficient	$\frac{2\sum_{j=1}^{d}\min(x_j, y_j)}{1-\sum_{i=1}^{d}(x_j + y_j)}$

Generally, clustering methods are divided into crisp and fuzzy. Crisp algorithms measure the data as each data is allocated exactly to one cluster. In some cases, we cannot allocate each data exactly to one cluster as some data are between clusters. The existing data in various clusters border are not obliged to be fully belonging to one of clusters and with membership degree ranging 0-1 show their relative membership. Fuzzy clustering methods are suitable tools to show real structure of these data. In this technique, each object with membership degree belongs to clusters. Thus, one object can be the member of two or some clusters at the same time. The higher the membership degree, the higher the membership of the object to that cluster.

Crisp methods are divided into partitional and hierarchical. Fuzzy clustering methods are extended form of crisp clustering methods. There are various fuzzy methods, each of them uses membership and objective functions. K-mean fuzzy methods, k-mode fuzzy and c-mean fuzzy are these methods.

One of the highly applied fuzzy clustering methods is c-mean fuzzy. This algorithm is developed form of k-mean method. In c-mean Fuzzy algorithm, numbers of Clusters (C) are determined already. Various methods are presented to determine the optimum number of clusters. Profile chart is one of the methods. Profile chart is based on non-similarity matrix and is a tool for quality of clustering. An index called Silhouette is presented by Rousseeuw and Silhouttes (1987) as referred in the followings. If object i is in cluster A, Silhouette index of object i is denoted by S(i) and is computed as followings:

$$S(i) = \frac{b(i)-a(i)}{Max\{a(i), b(i)\}}$$

where, a(i) is average distance of object i from all objects of cluster A(i) = $\min_c d(i, c)$ c-mean fuzzy algorithm attempts to minimize the following objective function:

$$J = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m} d_{ik}^{2} = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m}\left\|x_k - v_i\right\|^2$$

In the above equation, m is a real number bigger than 1 and in most cases, value 2 is selected for m. x_k is kth sample and v_i is center of cluster ith. U_{ik} shows membership to sample ith in kth cluster. Sign ||*|| shows the similarity (distance) of sample from cluster center and we can use any function indicating the similarity of sample and cluster center. According to U_{ik}, we can define a U matrix with c rows and n columns and its components are ranging 0,1. Matrix U components can take 0-1 but total of components of each of columns is equal to 1 as:

$$\sum_{i=1}^{c} u_{ik} = 1,\ \forall k = 1, ..., n$$

The meaning of this condition is that the sum of membership of each sample to c clusters is equal to 1. By above condition and minimizing objective function we have:

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^{m}\, x_k}{\sum_{k=1}^{n} u_{ik}^{m}},\ u_{ik} = \frac{1}{\sum_{j=1}^{c}\left(\frac{d_{ik}}{d_{jk}}\right)^{2/(m-1)}}$$

Thus, algorithm stages include:

- For m, c, suitable values are selected and U of initial clusters is guessed
- The clusters centers are computed (calculation of V is)
- Computation of membership matrix from the calculated clusters in 2
- If ||≤εUl+1-Ul||, algorithm is finished, otherwise we return to stage 2

Literature review: A study has been conducted by Nakhayias superiority of c-mean clustering in defining distribution of hydro chemical facies of water and underground system. In this study, clustering of a set of hydro chemical data >90 water samples and 9 hydro chemical variables) of Varamin plain is performed by fuzzy c-mean clustering methods and hierarchy cluster analysis and their application in interpretation of hydro chemical facies is discussed.

Babai *et al.* (2013) performed a study as the model of using data mining techniques in identification, division and analysis of behavior of customers of e-Banking services. His proposed model was based on CRISP standard process in data mining and after

preparation and pre-processing of data, two approaches are raised. Division of customers by clustering and computation of value of each customer in clusters and their ranking to find the most valuable clusters. Scoring and determining customer value as the goal feature in making classification models of customers' value. Davari conducted a study as increasing efficiency of k-mean clustering process by Hierarchy Methods. In this study, three hybrid cluster algorithms as using clustering process separately for determining the similar regions are investigated. Agglomerative hierarchical clustering algorithms, single linkage, complete linkage and ward and partitional clustering algorithm, k-mean algorithm were applied. A study has been done by Cheng and Chen (2009) as classification of customer value by RFM Model and RS theory. In this study, RFM Model is used to achieve quantitative values and then k-means algorithm for clustering customers' value (identification of loyal customers).

A study has been conducted by Chin as clustering and selection of providers based on simulated Anyl algorithm. This study proposed two optimization math models for clustering and selection of suppliers. The first model as the analysis of supplier cluster based on customer demand features as production cost, product quality and production time. Second model is achieved by source cluster in first model to determine suitable composition of suppliers.

The present study presents a comprehensive frame work for clustering customers of Golestan Company by considering four demographic variables. By fuzzy clustering approach and the data collected in the past about customers, consumers divided company products into groups with similar needs and features as customers inside each cluster have highest similarity in terms of variables in division and the highest distinction between clusters is occurred that in various stages of CRM, suitable marketing strategies are formulated for retention and absorption of each group of customers.

MATERIALS AND METHODS

The applied studies are those using theories, regulations and techniques as formulated for basic researches to solve executive and real problems. This study is applied in terms of purpose and is descriptive-analytic in terms of data collection. The study population is including all customers of Golestan Company in Kerman Province. Informatic expert of Golestan Company in central office has collected the statistics of transaction of customers during the mentioned period from data basis of customers in the form of Excel file. According to the collected statistics, totally 1591 customers purchased products of Golestan Company during 12 months in Kerman Province and purchase transaction of customer was incomplete and it didn't enter system. One of the customers returned their entire purchase. Thus, mentioned customers were eliminated from the customers list and finally, 1589 customers were studied in this study. Under real conditions, fuzzy clustering was much natural than hard clustering. To be familiar with theoretical issues and theoretical basics, library studies are applied and for data analysis, Excel, Matlab and SPSS are used.

RESULTS AND DISCUSSION

For problem solving and determining optimal number of clusters, at first numerical values are standardized with one of the six standardization methods. Then, optimal number of clusters is defined. To do this, standardized values enter Matlab Software in each method and optimal number of clusters is computed separately for entire customers, non-cash customers and cash customers. Silhouette coefficient is computed with the number of different clusters of minimum 2 clusters above. The results are shown in Table 4.

As Silhouette coefficient in c = 2 has the highest value and by increasing the number of clusters its value is reduced we stop clustering for higher values. The best number of clusters for total customers is 2. In this stage, by c-mean fuzzy method we determine mean membership degree. Due to high information volume, only membership degree of 20 people is shown in Table 5.

Membership degree of people shows that each person with which percentage is in each of clusters. The higher the membership degree percentage in a cluster, the higher the membership to that cluster. For example, first person with membership degree 0.4248 is in cluster one and with membership degree 0.5752 is in cluster 2 and as membership degree in cluster 2 is higher than that of cluster 1, we can say that person belongs to second cluster. The rest of people are in each of two clusters. Totally, of 1589 customers, 797 people are in first cluster and 792 people in second cluster and it shows high membership of people to cluster 1. Later, entire customers are divided into cash and non-cash customers and all stages for entire customers are repeated for these two groups.

Table 4: Results of computation of silhouette coefficient for standardized values

Silhouette coefficient	Values
C = 2	0/821*
C = 3	0/5074
C = 4	0/4537
C = 5	0/4444
Best choice	C = 2

*This value is determined as approximation

Table 5: Final membership degree of entire customers

	Object									
Clusters	1	2	3	4	5	6	7	8	9	10
1	0/5251	0/3967	0/4597	0/6007	0/5247	0/6182	0/5477	0/5096	0/4712	0/4517
2	0/4749	0/6033	0/5403	0/3993	0/4753	0/3818	0/4523	0/4904	0/5288	0/5483
	Object									
Clusters	1580	1581	1582	1583	1584	1585	1586	1587	1588	1589
1	0/4248	0/4372	0/4823	0/5566	0/4699	0/5592	0/4393	0/5927	0/4860	0/5070
2	0/5752	0/5628	0/5177	0/4434	0/5301	0/4408	0/5607	0/4073	0/5140	0/4930

Table 6: Results of computation of Silhouette coefficient for customers with non-cash purchase

Silhouette coefficient	Values
C = 2	0/783*
C = 3	0/4298
C = 4	3367
C = 5	0/3623
Best choice	C = 2

*It is determined as approximate

Clustering customers with non-cash purchase: In definition of non-cash customers there is a group, their purchase is done via check. The computation criteria for customers with non-cash purchase are based on orders in 2013, annual purchase price (in Rial) and the number of their bounced checks in 2013. The results of optimal number of cluster for customers with non-cash purchase with computation of Silhouette coefficient by Matlab Software are shown in Table 6.

As shown in Table 7 the values for Silhouette coefficient are the best number of clusters for customers with non-cash purchase as 2. Table 7 shows the summary of results of computation of membership degree of customers with non-cash purchase (due to great number of samples, only 10 first and last samples of customers are shown). Totally, 816 people are in cluster 1 and 773 people in cluster 2.

Clustering customers with cash purchase: In definition, cash customers are those their purchase is in cash. The computation criteria for customers with cash purchase are based on number of orders in 2013 and annual purchase price (in Rial) (Table 8).

Thus, the best number of clusters for customers with cash purchase is 2. Clustering cash customers is shown in Fig. 2 and their membership degree is shown in Table 9.

As shown in Table 1-12, total people are divided into two clusters and 796 people are in cluster 1 and 793 people in cluster 2 and it shows that number of cluster 1 is higher than that of cluster 2.

By investigation of membership degree of three divisions (entire customers, cash customers and non-cash customers) we find that number of cluster 1 people is more

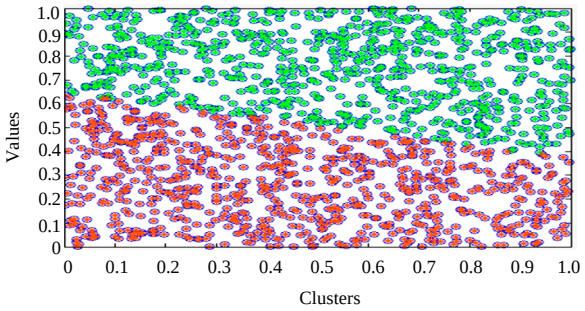

Fig. 2: Clusteirng cash customers

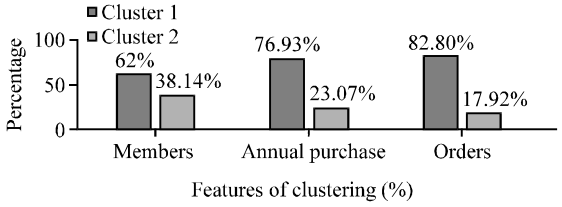

Fig. 3: General features of clustering customers

than that of cluster 2. It means that people with high membership degree belong to first cluster. In second phase of study by SPSS Software and k-mean clustering method, we define statistical feature of each cluster to the indices. The summary of the results is shown in Table 10.

As shown in Table 10, all customers of cluster 1 performed their purchases with presenting at least one check but all customers of cluster 2 purchase in cash. Not only the number of customers of cluster 1 is higher than that of cluster 2 (it means that many customers are inclined to non-cash purchase and check purchase) and the purchase volume of these customers is higher than that of the customers purchasing in cash. Also, the number of annual orders of cluster 1 customers is higher than that of cluster 2 customers. The results are shown in Fig. 3. Statistical features of customers with non-cash purchase are shown in Table 11.

As shown in Table 11, second cluster customers are higher. They have high orders in this company. The

Table 7: Membership degree of customers with non-cash purchase

	Object									
Clusters	1	2	3	4	5	6	7	8	9	10
1	0/4651	0/3393	0/1439	0/7094	0/4887	0/6659	0/2864	0/6710	0/3324	0/6427
2	0/5349	0/6607	0/8561	0/2906	0/5113	0/3341	0/7136	0/3290	0/6676	0/3573
	Object									
Clusters	1580	1581	1582	1583	1584	1585	1586	1587	1588	1589
1	0/3406	0/3666	0/3208	0/2471	0/4168	0/5498	0/3430	0/7507	0/8178	0/2951
2	0/6594	0/6334	0/6792	0/7529	0/5832	0/4502	0/6570	0/2493	0/1822	0/7049

Table 8: Results of computation of Silhouette coefficient for customers with cash purchase

Silhouette coefficient	Values
C = 2	0./762*
C = 3	0/6480
C = 4	0/4383
C = 5	0/4111
Best choice	C = 2

*It is approximate

Table 9: Membershp degree of clusters of customers with cash purchase

	Object									
Clusters	1	2	3	4	5	6	7	8	9	10
1	0/1321	0/3745	0/4837	0/9232	0/0848	0/5346	0/9255	0/1493	0/1279	0/6013
2	0/1984	0/2293	0/4692	0/8663	0/3626	0/9108	0/1747	0/4238	0/2609	0/8945
	Object									
Clusters	1580	1581	1582	1583	1584	1585	1586	1587	1588	1589
1	0/8016	0/7707	0/5308	0/1837	0/6374	0/0892	0/8253	0/5762	0/7391	0/1055
2	0/8679	0/6255	0/5163	0/0768	0/9152	0/4654	0/0745	0/8507	0/8721	0/3987

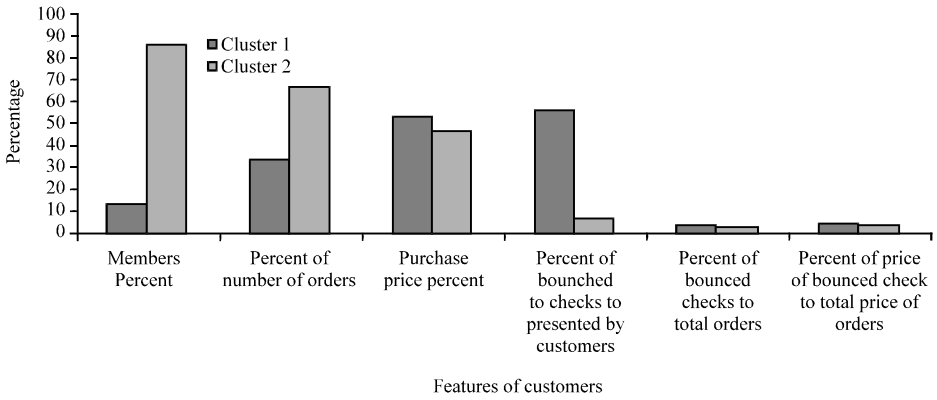

Fig. 4: General features of customers with non-cash purchase

number of their orders to the number of cluster members is lower than that of first cluster. Volume of orders is not different considerably from the first cluster in Rial and is less compared to the number of members from first cluster. The customers of second cluster presented many checks to company but their bounced checks are low in terms of number, Rial and compared to the number of cluster members and number of orders compared to that of first cluster. The results are shown in Fig. 4.

Statistical features of customers with cash purchase are shown in Table 12. Results of customers clustering with cash purchase show that only two people are in cluster 1 and the number of cluster 2 members is more than that of cluster 1 members. They have higher orders compared to them and annual Rial purchase volume of cluster 2 members is higher than that of cluster 1. The results are shown in Fig. 5.

Table 10: Statistical features of clusters for entire customers

Indices	Cluster 1	Cluster 2
Members		
N	983	606
%	61.86%	38.14%
Number of orders		
Maximum	98	57
Minimum	1	1
Mean	15/12	5.91
Sum	11944	3581
%	76.93%	23.07%
Annual purchase price (Million Rial)		
Maximum	2967.77	230.24
Minimum	1.07	0.14
Mean	134.66	47.68
Sum	132374	28893
%	82.08%	17.92%
Cash/check purchases		
Number of purchases with check	983	0
Number of cash purchases	0	606
Bounced checks		
Maximum	12	-
Minimum	0	-
Mean	0.82	-

Table 11: Statistical features of customers' clusters with non-cash purchase

Indices	Cluster 1	Cluster 2
Members		
N	134	849
%	13/63%	86/37%
Number of orders		
N	986.3	958.7
%	33/37%	66/63%
Mean	29/75	9/37
Max.	98	34
Min.	5	1
Annual purchase price (Million Rial)		
Total annual purchase	085.70	289.62
Purchase percent	52/94%	47/06%
Mean	523.02	73.37
Max.	967.77.2	645.24
Min.	25.4	1.07
Bounced check		
Total checks (N)	787.1	4,054
Total bounced checks	425	380
Percent of bounced checks to presented checks by cluster customers	23/78%	9/37%
Percent of bounced checks to total presented checks by customers	7/28%	6.51%
Mean	3.17	0.45
Max.	12	4
Min.	0	0
Percent of bounced checks to the number of orders of cluster	10.66%	4/78%
Percent of bounced checks to total orders	3/56%	3/18%
Total price of bounced checks (Price (Million Rial))	7,013	4,711
Percent of bounced checks to total price of cluster orders	10.01%	7/56%
Percent of price of bounced total price checks to of orders	5.30%	3/56%
Mean	62.06	17/45
Max.	238.03	142
Min.	0.62	0.9

Table 12: Statistical features of customers' clusters with cash purchase for clustering with two clusters

Indices	Cluster 1	Cluster 2
Members		
N	2	604
%	0.33%	99/67%
Number of orders		
N	77	3,504
%	2/15%	97/85%
Mean	38.5	5.8
Max.	57	48
Min.	20	1
Annual purchase price (Million Rial)		
Total annual purchase	3,701	25,192
Purchase percent	12.81%	87.19%
Mean	1850.50	41.71
Max.	2,302	941.75
Min.	1.398	0.14

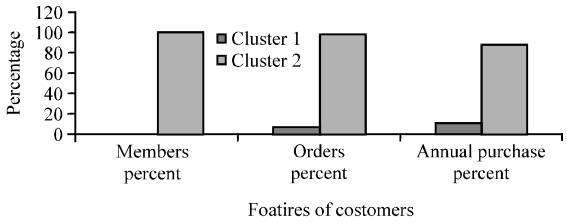

Fig. 5: General features of customers with cash purchase

CONCLUSION

By using data mining approaches, companies perceive hidden communication in data, behavioral models and requirements of customers and they can dedicate their resources better to fulfilling the needs of customers, presenting new services and accepting customers. Based on great amount of data of customers of Golestan Company and lack of appropriate use of managers and experts of these data, researchers attempted to divided them into homogenous groups based on various and criteria as measured for each customer and to identify and analyze the behavior of customers and presenting special services to various groups.

By Matlab Software, optimal number of clusters for all groups is 2. The membership degree of each of customers is determined based on number of clusters and in three classifications, the number of first cluster is higher than that of second cluster and customers were in second cluster with higher membership degree. Then, statistical feature of customers was determined by SPSS Software. Totally, the highest annual purchase was dedicated to customers of cluster 1 as 132374 Rial and it showed that totally 82% of total purchases in 2013 were dedicated to this group. The number of orders of

cluster 1 was higher than that of cluster 2. The customers of cluster 2 performed their purchases as cash and customers of cluster 1 performed their purchase via check and they were higher. Thus, purchase volume of cluster 1 was higher than that of cluster 2. Thus, company dedicated much time and cost to keep and increase satisfaction of customers of cluster 1 with cash purchase.

As cluster 1 with cash purchase had high purchase volume for each person, by dedicating low costs and special facilities, their satisfaction is increased and these facilities include as:

- Agency with the lowest cost: reduction of trading costs and presenting product and services
- Proposing value to customer: rapid execution of services, easy access to products and services by e-Business
- Customization: customized services and products are formulated based on taste of customers

As customers of cluster 1 with non-cash purchase had the highest number and price of bounced checks, using bonus as discount in future purchases in case of timely payment of obligations and no bounced check can be proposed. Also, we can use inhibiting policies (e.g., lack of purchase permission for customers not liquidating >21% of their checks).

Customers of cluster 2 with non-cash purchase have about 70% of number of customers and orders and only about 40% of annual purchase volume is used. By suitable encouraging policies, we can encourage them to higher purchase and increase trading volume. The following solutions are proposed:

Customer orientation culture: Assuring the customers as all parts of company by giving continuous training regarding products and services have the highest quality, concentration and commitment to customers.

The company can compute loyalty of customers of each cluster and take marketing strategies and CRM and implement them and creates loyalty between customers of clusters with low loyalty.

This study clustered a part of customers of Golestan Company. This study is proposed to total customers. We can also use other methods for division of customers. Clustering with non-financial indices as purchase variety and weighting for each index based on importance in decision making can be used for further studies.

REFERENCES

Babai, Kh.M., E.l. Zine and F. Abidine, 2013. Model using data mining techniques in the identification, segmentation and customer behavior analysis electronic banking services. New Market. Res. J., 2: 175-188.

Berson, A., S. Smith and K. Thearling, 2000. Building Data Mining Applications for CRM. Tata McGraw-Hill Publishing Company Ltd., New Delhi.

Cheng, C.H. and Y.S. Chen, 2009. Classifying the segmentation of customer value via RFM model and RS theory. Expert Syst. Appl., 36: 4176-4184.

Dastgir, M.S. and M. Sardasht, 2011. Data mining technology, new approach in financial field. Accounting Knowl., 5: 6-27.

Hsu, C.C., V.R. Kannan, G.K. Leong and K.C. Ten, 2006. Supplier selection construct: Instrument development and validation. Int. J. Logistic Manage., 17: 213-239.

Jamal, A. and K. Anastasiadou, 2009. Investigating the effects of service quality dimensions and expertise on loyalty. Eur. J. Marketing, 43: 398-420.

Rousseeuw, J. and P. Silhouttes, 1987. A graphical aid to the interpretation and validation of cluster analysis. J. Comput. Applied Math., 20: 53-65.

Rygielski, C., J.C. Wang and D.C. Yen, 2002. Data mining techniques for customer relationship management. Technol. Soc., 24: 483-502.