IJEP International Journal of New Political Economy

DOI: 10. 52547/jep.3.1.129

Customer Clustering Based on RFM- CPI Model Using Data Mining Techniques

Abdolmajid Imani[1], Meysam Abbasi[2], Farahnaz Ahang[3], Hassan Ghaffari[4]

ARTICLE INFO

Article history:

Date of submission: 21-04-2021

Date of acceptance: 16-04-2022

JEL Classification:

C80

H57

L16

Keywords:

Data mining

RFM model

Customer Price Index (CPI)

ABSTRACT

Subject and purpose of the paper: the goal of this article is suggesting a new method for increasing the quality of customers clustering with increasing Customer Price Index (CPI) to monetary variable (M). At this research, customers of one chain store of Zahedan, Iran according to RFM-CPI and RFM basic model variables and Two-step algorithms are clustered to compare these two procedures. Furthermore, determining the best method for customer clustering can be done. At this research, different steps of data mining and data analysis for discovering the knowledge of them were done according to the standard process of CRISP-DM (1); this process includes System understanding, Data understanding, data preparation, Modeling, Model assessment and deployment. According to the results, increasing the Silhouette index at the RFM-CPI model in recent article in comparing with the basic RFM model defines high accuracy.This corrected model has advantages toward the main model; these advantages are contained: monetary changes at a period of time are identified, also according to clustering.

1. Introduction

Different strategies may be followed by different segments. Some changes may be happened by sheering from production-based business strategy and

1. Assistant Professor, Faculty of Management & Economic, University of Sistan and Baluchestan, Zahedan, Iran
2. Master of IT Management, Faculty of Management & Economic, University of Sistan and Baluchestan, Zahedan, Iran
3. Ph.D. Student of Public Administration, Faculty of Management & Economic, University of Sistan and Baluchestan, Zahedan, Iran, (Corresponding Author) Email: ahang1989@yahoo.com.
4. Ph.D. Student of Public Administration, Faculty of Management & Economic, University of Sistan and Baluchestan, Zahedan, Iran

joining to concentrate on customer strategy. Nowadays, competing is an important issue in the business world and for being successful in this part, not only paying attention to attract new customers is enough, but also maintaining old customers is a more important issue, and they should believe that by attending to that organization some profits may be created and kept for them (Ghalenooie & Sarvestani, 2016, 365). One of the most important strategies for maintaining customers is customer relationship management, and nowadays it can be mentioned this strategy can be known as the most important strategy in the business world (Soltani, Zareie, Milani, & Navimipour, 2018, 668). Some advantages like Web/App technology makes customer relationship management more important than the other's issue because it makes them more capable of understanding the customers and their potential, so they may be convinced to make transactions and decisions (Anshari, Almunawar, Lim, & Al-Mudimigh, 2018, 2). Different ideas and attitudes are mentioned by different firms and experts, so there is no identical definition of CRM. Lager (2008) considered CRM for technology. With this technology, a firm or company may be enabled to have high sales. Others conjoin CRM with processes of data management, in a way that better profits can be reached and connections with higher firmed can be created (Barson et al., 2000). Other authors considered CRM as a process that provides the highest advantages on the double side (Zerbino, Aloini, Dulmin, & Mininno, 2018, 4). Explained CRM is explained as a strategic issue that integrated all processes in a way that high values for customers and firms can be created (Buttle, 2009, 2; Greenberg, 2010, 5; Triznova, Mat'ova, Dvoracek, & Sadek, 2015, 954). Some chances are provided by CRM, which can use data and information to perceive customers and create high values (Haislip & Richardson, 2017, 16). Performing CRM is an important step in any CRM strategy so by reaching the success of this part high values and improvements can be created (Li, Huang, & Song, 2019, 2). Salesforce automation, data warehousing, decision support, reporting tools, and data mining are some examples of CRM system functionality (Talón-Ballestero, González-Serrano, Soguero-Ruiz, Muñoz-Romero, & Rojo-

Álvarez, 2018, 187). Data mining is defined as the process that applies statistical, mathematical, and artificial ingenuity and machine learning techniques for useful information identification and extraction to bring knowledge from databases (Ngai, Xiu, & Chau, 2009, 2593). Extracted information from data mining in CRM showed that all customers participate in businesses uniformly, and it is necessary to mention, customer clustering before designing can create maximum profits (Hu, Huang, & Kao, 2013, 779). Customer clustering is created by using behavioral data because it is commonly available and includes the time and procurement history (Harwati & Karunia, 2017, 6407). A technique that is used for customer evaluation based on their procurement manner, is named RFM analysis (Christy, Umamakeswari, Priyatharsini, & Neyaa, 2018, 1). One of the famous and powerful instruments in database marketing is RFM analysis, which is extensively applied customer values measured according to their previous procurement history (Tanaka, Hamaguchi, Saigo, & Tsuda, 2017, 957). A lower value is related to a higher probability of repeating the procurement by customers. The number of procurement in a period of time is the frequency and shows the loyalty of customers. The money which is spent over a period of time is known as monetary (Mohammadzadeh, Hoseini, & Derafshi, 2017, 24). One of the economical indexes which can measure money changes in a certain period of time is the Customer price (CPI) index. CPI is a measurement of price level changes of customer products and services in a special period of time. CPI is closely related to the national economy and people's lifestyles, and it is an important commission for managing macroeconomic politics and money market operation analysis, bond market and central bank's open market, and the rate of its change, which reflects inflation degree or destruction to a certain gamut. Many factors like the policy of monetary, industrial diplomacy, and investment demands are influencing the CPI oscillation (Xiao, Wang, Tian, & Zhen, 2018, 3). According to the mentioned contents, analyzing the results by using of CPI index in a monetary variable at RFM- CPI model than RFM basic model for determining a better model and qualified clustering is the goal of recent

research. In future sections, some basic concepts and previous research will be reviewed and the suggested framework will be presented.

The present article is organized in five sections. The second part explains the theoretical. In the third part, the methodology. In the fourth section, Findings. In the fifth section, Conclusion and recommendations

2. Theoretical Framework

2.1 Clustering

One of the most important titles in data mining is clustering, which is called cluster analysis. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters (Rajeh, Koudehi, Seyedhosseini, & Farazmand, 2014, 121). Clustering is a process that includes different segments and in each segment, members with the highest similarities have existed (Wang et al., 2018, 245). About each cluster, there is a head (CH) which is a leader of the cluster, and the members (Ms) of the cluster are influenced by the head and aggregated data are led to the base station (Gupta, 2018, 235). In the literature, different behaviors exist for data clustering (segmentation), and two principles are hierarchical and partitioning clustering. About the hierarchical clustering, the clusters are inside each other. When the data are separated into different levels, this clustering category is used (Chemchem & Drias, 2015, 1437). In the marketing literature, categorizing the customers is done by using clustering-based segmentation to put the customers in the clusters according to their benefits and product attribution (Casidy, 2012, 7). A necessary level for the complexity reduction of large-scale logistics network optimization is customer clustering. By optimal and suitable clustering, the customer with identical traits can reduce operational costs and improve the levels of customer satisfaction (Wang, Ma, Lao, & Wang, 2014, 1). For identifying different types of customers and segment them according to their similar profits, some points like terms of behavior requirements want and characteristics and the main aim of clustering techniques are important

(Tripathi, Bhardwaj, & Poovammal, 2018, 803). For detecting the varieties among the customer's attitudes and intentions, observable shopping behavior can be applied (Bose & Chen, 2015, 3). Different clustering algorithms for clustering around the customers according to their traits have existed. In most the clustering algorithm, applied priority for measuring the result's clusters quality is explained as in Eq1, which is known as minimizing the sum of squared error.

$$f(O,C) = \sum_{i=1}^{k} \sum_{O_i \in Z_i} d(O_i, Z_i)^2 \quad (1)$$

Where d (Oi, ZI) determines the distance between an object Oi and the cluster centroid ZI. Similarity and dissimilarity between objects are defined through some distance functions. The most usual distance function is the Euclidean distance that is expressed as follows:

$$d(O_i - O_j) = \sqrt{\sum_{p=1}^{d} (O_i^p - O_j^p)^2} \quad (2)$$

Where d (Oi, Oj) determines the distance between two objects Oi and Oj, P is dataset girth. K-means, graph, hierarchy, and Two-step are clustering around the algorithm. In recent research Two-step clustering algorithm is applied according to the data nature used and the problem is related to the classical clustering method. An abbreviated definition of the mentioned method is defined in the next section.

2.2 Two-step cluster analysis

There are different amounts of datasets and for measuring each of them a special algorithm must be used. For managing large datasets, Two-step cluster analysis algorithms are used (G. Li & Sun, 2018, 3), and an important point about the hierarchical and K-means algorithm is, that when n is very large these algorithms don't scale efficiently (Garson, 2009). For managing the variables with 3 or more clusters this algorithm is applied and can be used for both categorical and continuous variables (Zhang, Ramakrishnan, & Livny,

1996, 103). When a hypothesis isn't as reasonable as people thought, this algorithm can make the results in a logical way. This cluster analysis is more applied than the other cluster analysis algorithms, and this claim is because of testing the assumptions and calculating the observed data, and this algorithm is more accepted than the others because of changing the assumptions for best performance (Tekień, Gutkowska, Żakowska-Biemans, Jóźwik, & Krotki, 2018, 175). The two-step clustering algorithm has two steps as it is indicated in its name: in the first step K-means algorithm is applied for clustering around the data and the K-means algorithm is as follows:

Selecting the population and distance between the individual is an important segment. For the primary centers, K random individual is chosen. Calculating individual distances from all centers and collecting them in the closest center are essential points to be considered. Taking the means of each cluster is done to reduce the changes in each cluster and by doing this function, calculating new cluster centers can be performed. The hierarchical agglomerative clustering method is applied in the next step, and it is done according to the results of the previous step. In this step, each member is collected into a cluster, and these clusters are agglutinated and created a unique cluster (Faezy razi & Shadloo, 2017, 51). The advantage of this procedure clustering algorithm about comparing with the K-means algorithm is that the results are dispassionate with primary parameters. The steps of hierarchical clustering algorithms are done as follows (Merceron & Yacef, 2004, 36). For the first segment, selecting the population and distances between individuals must be done. Each member is collected in a primary cluster. Calculating the distances between all clusters is done to create a distance matrix. While there is more than one cluster and the distance between two to the closest clusters falls below a given sill and collecting these two closest clusters into one cluster will be done. Calculating the distances between all other clusters will be done again, and a new cluster must be formed. In the recent article, IBM SPSS Modeler 18 was applied, and it was done for clustering with a Two-step algorithm.

2.3 Data Mining Techniques

The craft of detecting the information or knowledge in a large amount of data is known as Data Mining (DM). For example, in statistics, data mining has become usual in firms and organizations. That information extraction from their database can be done to be applied to their own requirements (Rémy, Martial, & Clémentin, 2018, 121). There is no standard terminology for DM technique, but algorithm and technique are the most cherished epithets. For expressing the wide range of methods, which are entrained in DM, 'technique' is a proper word, and for defining the details, algorithms can be a proper one. Anyway, they are both applied for an identical concept. According to Berry and Linoff DM techniques contained: analysis of market-based, the reasoning memory-based, cluster discovery, link analysis, decision trees, and rule induction, artificial neural networks, algorithms of genetics, and processing of online analytics (Aghimien, Aigbavboa, & Oke, 2019, 3242). According to the working concept, every data mining is distributed into two large segments. According to storage or distilled for data usage, the higher level of this clustering is assigned:

1 .Using the data in a direct way or saving them. In this part, the original data is kept in detail and apparently applied for forecasted modeling steps and/or analysis exceptions. Analyzing by this procedure may be difficult in very large databases, so it can be an obstacle for this method. The procedures like the nearest neighbor and K-nearest neighbor have appertained to this segment.

2 .Formal law's distillation templates usage and identification. When distillation technology extracted one sample, extracted information can be available from raw data, which is transformed into a certain formal structure that relies on the data mining procedure. The mentioned procedure is run at the level of free search. An important point in this part is the absence of the first segment of the procedure (Kochetov, 2018, 237).

2.4 RFM Model

One of the simplest and the most powerful model for CRM is the RFM model (Carrasco, Blasco, & Herrera-Viedma, 2015, 1341). In CRM, quantitative traits are discovered by the RFM model, and the variable is enriched for 15 potential relationships because the customer value can be inverted by the most maiden consumption as recency, frequency, and monetary in the model (Song, Zhao, E, & Ou, 2017, 9). The RFM details are expressed as follows:

1. The recency of the last transaction which is observed by (R) refers to the period of time between the last transaction time and a certain Period. A lower interval shows a higher value of this variable in the model.
2. Frequency, which is observed by (F) means transactions number at a certain period of time. When the purchase repetition is higher.
3. Monetary, which is observed by (M) refers to the value of money consumption by the customer in a certain period of time (Cheng & Chen, 2009, 4178). The RFM model has been applied in industry and direct marketing for more than 30 years, mostly due to its simplicity. This model is grounded on the analysis of customers' past behavior and assumes that those with a desirable value for each of the model's indices are the best customers as long as their future behavior is the same as the past (Alizadeh Zoeram & Karimi Mazidi, 2018, 365). The process to quantify customer behavior via the RFM model is as follows. First, sort the database by each dimension of RFM and then divide the customer list into five equal segments. The method is known to have an exactly equal size. Different RFM quintiles have different response rates. For recency, customers are sorted by purchase dates. For frequency, the database is sorted by purchase frequency (the number of purchases) made in a certain time period. For monetary, customers are coded by the total amount of money spent during a specified period of time. Definition of money is defined by the dollar value that the customer spent in this time period or

by the average dollar amount per purchase or all purchases to date (Mohammadian & Makhani, 2016).

Customer Price Index (CPI): Price and price opinions play an important role in the science of the economy. The price is known as an economic phenomenon, and it can be derived from Aristotle and Xenophon's explanation. The price is the amount of money, which must be paid from a buyer to the seller for a special product (Dolca & Nicolov, 2013, 739). A measurement for recognizing the changes in the prices of merchandise's basket is known as index numbers, which have been in the official application for a century and a half (Chance, 1966, 108). One of the indexes for national statistical offices that are used for different target varieties is known as Laspeyres and Paasche gnomon. Customer Price Index (CPI) is defined as the most frequently applied official indicator in many countries that asserts to measure the changes in product costs; so they are measured by the product's basket of the contingent group in a special period of time (ILO, 2004). CPI is a kind of indicator for measuring the average prices for those kinds of products, which are bought and used by homebred. Inflation is defined as the percentage of changes in CPI (Burns, Sacks, & Gold, 2008, 450).

2.5 Experimental history

Chou and Shu-Chen analyzed The RFM Mode for VIP Customer. this study proposed customer ranking combined big data process based on the RFM model (recency, frequency, monetary) to develop a recommendation algorithm using an association rule, which finds greater recommendation to promote operational effects of firms. The authors adjust the weight of potential information to perform the customer ranking, which is conducted by using agglomerate hierarchical clustering (Chou and Shu-Chen, 2020).

Hu et al classified Electric vehicle user based on charging big data. Analysis results demonstrate that 7426 EV users are divided into six groups, namely “high value users”, “key users to maintain”, “key users to develop”, “potential users”, “new users” and “lost users (Hu, Zhou, Li &Ma, 2022).

Doğan et al segmented customers by using RFM model and clustering methods. They propose two different clustering models to segment 700032 customers by considering their RFM values. They suggest that the current customer segmentation which built by just considering customers' expense is not sufficient. Hence, one of the models that recommended in this research is expected to provide better customer understanding, well-designed strategies, and more efficient decisions (Doğan, Ayçin & Bulut, 2018).

Safari et al analyzed the customers by using of fuzzy analytic hierarchy process and fuzzy clustering. They divided customers into 9 clusters by using of FCM algorithm; at the end, they ranked the customers by calculating the CLV of each cluster (Safari, Safari, & Montazer, 2016). Peker et al, classified the customers of the food stores in Turkey. They analyzed the customers by using of LRFMP model and K-means algorithm. Results showed five clusters (Peker, Kocyigit, & Eren, 2017). Wu Cheng and Lo, analyzed the customer value of one industrial type of equipment company by using of RFM model and K-means clustering. They divided the customers into six clusters and analyzed them by using CLV. Finally, appropriate strategies for different clusters are suggested (Wu, Chang, & Lo, 2009). Chang and Tsay, classified the customers of one store by combining issues of SOM and K-means and using of LRFM model. Results caused the creation of five kinds and 16 clusters of customers. Also, results showed that increasing variable L caused improved customer identification (Chang & Tsay, 2004). Imani and Abbasi, analyze the customers of one chain store in Iran; it was done by using of RFM model. They applied the fuzzy analytic hierarchy process and fuzzy C-means algorithm; they divided the customers into seven clusters and provided a background for codifying relationship strategies with the customer (Imani & Abbasi, 2017). Mozaffari et al, did research with the goal of clustering and exploring customer damage patterns of tertiary person insurance for identifying the customers by using the index of Davies Bouldin and clustered around the customers with the K-means algorithm. Finally, the customer value of each cluster was calculated and

clusters were ranked according to their values (Mozaffari & Bonyadi, 2010). Schiopu, analyzed the customer profiles of the bank. He clustered around the customers by using of Two-step algorithm. Research results caused improved company profit with the management of customers (Şchiopu, 2010).

3. Methodology

In this research, customers of one chain store in Zahedan, Iran according to RFM-CPI and RFM basic model variables and Two-step algorithms are clustered to compare these two procedures. Furthermore, determining the best method for customer clustering can be done. In this research, different steps of data mining and data analysis for discovering their knowledge of them were done according to the standard process of CRISP-DM (1); this process includes System understanding, Data understanding, data preparation, Modeling, and Model assessment and deployment. Different Phases of this process are as figure 1.

4. Findings and Conclusion

4.1 Phase 1: System understanding

Step 1: Goals understanding and Firm situation: Refah Chain store is known as one of the most widespread recent product distribution networks with the goal of preparation, supply, distribution, and sale of essential products. This company makes its products available to its customers and by holding discounts, and seasonal festivals try to do this importance as well as possible; in the same way, because of transactions growth in companies and applying them by data mining science experts, customer value identification can be useful for these companies' managers. Analyzing the results by using of CPI index in M variable at the RFM model than RFM basic variable is the main goal of this research so qualified clustering can be determined.

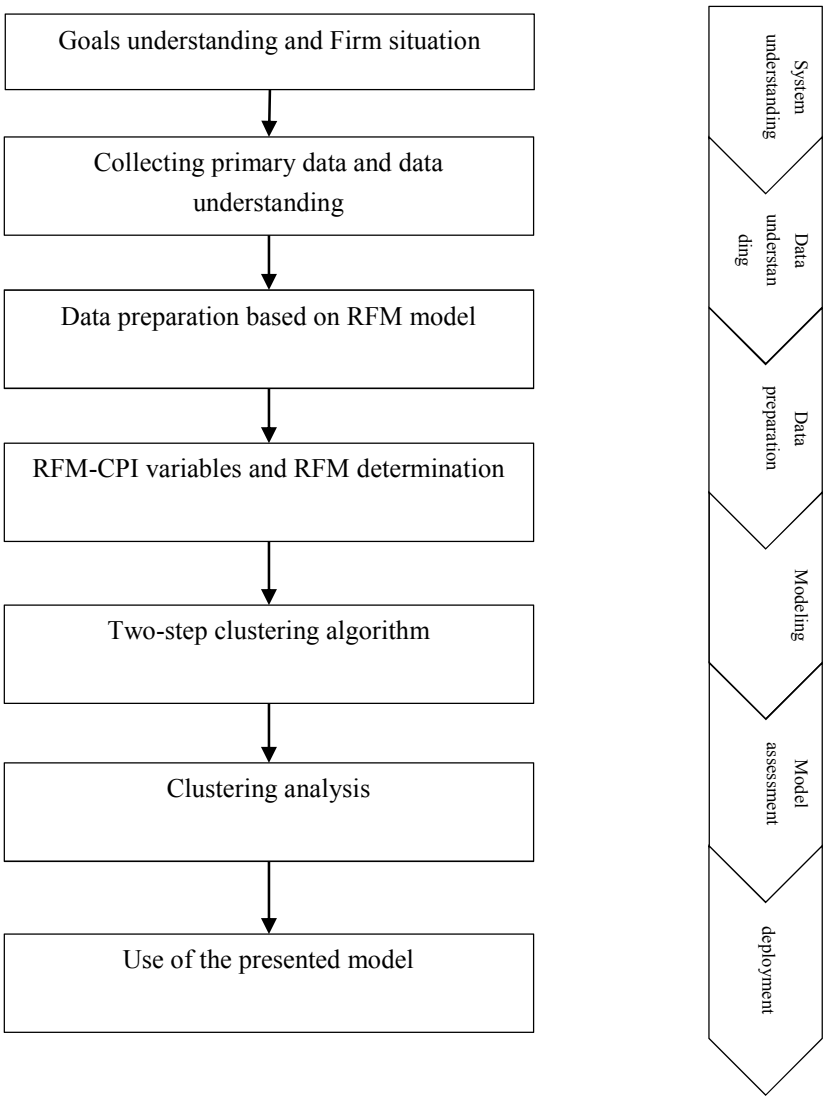

Figure 1: Research Phases, Source: Compiled by Authors

4.2 Phase 2: Data understanding

Step 2: Collecting primary data and data understanding: This step is started with the primary collection of data and then describes the data. At this level store, transactions are collected and a proportional database with research

requirements is formed. Data is collected from 22 September 2016 to 20 March 2017 and included 6665 customers. Collected data are as column and raw collection in excel file format, in a way that, columns have represented the traits like customer number, transaction date and the cost which is paid.

4.3 Phase 3: Data preparation

Step 3: Data preparation based on RFM model: The preparation data level contains all activities which are applied for the ultimate data set built (the data that are prepared for modeling) from primary raw data. If the quality of this preparation is better, modeling will be improved. For this purpose, at this step wrong and imperfect data, invalid amounts, and some transactions with lost amounts were omitted. Used variables in this research were: Recency (R), the interval from the last transaction of the customer to the end of the studied period (20 March 2017). Frequency (F): number of transactions from 22 September 2016 to 20 March 2017. Monetary(M): money volume of the customer at the indicated period. M variable is corrected according to monthly CPI index information for the year 2016-2017. Table 1 shows the CPI index amount at each one of the mentioned.

Table 1: Resource: Central bank of Iran (CBI)

CPI index amount	Month
100	23 September-22 October
100/2	23 October- 21 November
101/7	22 November- 21 December
102/5	22 December- 20 January
103/7	21 January – 19 February
105/9	20 February- 19 March

Source: Compiled by Authors

At the next step recency, frequency, monetary and monetary- CPI related to each customer is calculated by SPSS Modeler 18.

Step 4: RFM and RFM- CPI variables determination: By using formula 3, the monetary of the customer according to the CPI index amount based on table 1, was corrected. Table 2 shows the results of the corrected monetary amount besides M amount at the RFM basic model. For transforming the Rial value of money at the period of time toward another period of time, product cost index and consumption service (CPI) is used as formula.

Rial value of money= variable number at a special period of time/index number in the first, period of time*Rial amount

Table 2: An example of research data along with modifying the monetary variable by the CPI index

Customer number	Recency	Frequency	Monetary	Corrected monetary
001	20	18	26885204	27791903
002	4	60	51156449	51638593
003	32	10	36287622	37962034
004	51	4	20732369	21351414
. . .	. . .	. . .	. . .	. . .
6665	98	3	41840007	44222304

Source: Compiled by Authors

4.4 Phase 4: Modeling

Step 5: Two-step clustering algorithm: In recent research, a Two-step algorithm is used for customer clustering. For this purpose, SPSS Modeler 18 as powerful software for data mining was used. Clustering around quality assessment and excellence amount of one cluster toward other clusters are done by different clustering algorithms or identical algorithms but with different parameters amount, One of the indexes used to measure the quality of clustering is the Silhouette index. Silhouette index amount means is applied for clustering validity assessment and decision making about optimal cluster

selection, and this amount is calculated according to the distance and proximity of observations to each other. S(i) is calculated by below formula 3:

$$S(i) = \frac{(b(i)-a(i))}{\max(a(i).b(i))} \quad (3)$$

In this equation, an (i) represented the mean interval between observations, (i) with other observations at the same cluster, and b(i) represented the mean interval of observation, (i) to all observations at other clusters. According to the above formula S(i), the amount must be from -1 to +1. If S(i) is closer to +1, sample clustering was done well, and suggested cluster for the mentioned sample is appropriate and if S(i) is closer to -1, it means that sample clustering wasn't done well, and suggested cluster wasn't appropriate for intended data. Table 3 shows the results of clustering based on the Two-step algorithm in RFM basic model and RFM-CPI model. The details of the clustering with the chosen model are shown in Table 3.

Table 3: Silhouette amount based on RFM and RFM-CPI model

Used model	Cluster number	Silhouette amount
RFM	3	0.536
RFM-CPI	3	0.556

Source: Compiled by Authors

Table 4: RFM-CPI model details

Cluster	Average of recency	Average of frequency	Average of corrected monetary	Percentage	Number
1	21	5	4109729	69.5%	4629
2	77	4	3200066	20%	1333
3	16	14	14279353	10.5%	703

Source: Compiled by Authors

As, it is indicated at the tables 3 when RFM-CPI model is used for clustering, the Silhouette index amount is 0.556, and it shows a higher quality of clustering versus RFM basic model.

4.5 Phase 5: Model assessment

Step 7: Clustering analysis: Clustering analysis is performed from the comparison of RFM model variables average at each cluster and average of that index in all data and also variables ranking according to RFM model variables average. In this research, detail of RFM-CPI model clustering for clustering analysis was used, and it is presented in the format of table 5.

Table 5: Data clustering analysis

Cluster	Average of R	Average of F	Average of corrected M	Variables situation (R, F, M)	Variables rank (R, F, M)
1	21	5	4109729	(↑ ↓ ↓)	(2, 2, 2)
2	77	4	3200066	(↓ ↓ ↓)	(3, 3, 3)
3	16	14	14279353	(↑ ↑ ↑)	(1, 1, 1)
Total average	32	6	5000451		

Source: Compiled by Authors

In table 5, each one of the RFM-CPI model variables averages at each cluster is compared with its average for all data. In this table, this sigh (↑) represented the situation in which the variable average at one cluster is more than the value average of that variable in all data, and it shows an optimal situation of the variable. This sign (↓) represented that the variable average at one cluster is lower than its average in all data, and it shows an undesirable situation. As it is indicated, in the third cluster amount of all RFM variables is higher than the total average of data, and it shows an optimal situation for this cluster. In the following, customers are analyzed according to tables 4 and 5:

Cluster 1: This cluster contains 4629 members, and they form 69.5 percent of all chain store customers. From RFM- CPI model variables, they take place at second rank. The average of M variables is lower than the total average, but purchase recency is higher than average. The main goal of a store is to make effective relationships with them for leading to a loyal

cluster. So by codifying effective strategies of relationship with the customer, purchase number monetary volume can be increased.

Cluster 2: second cluster customers are 20 percent of all customers and according to cluster analysis, all RFM- CPI model variables are lower than the data average, and it shows an undesirable situation. From ranking, all variables took place at third rank; so for adopting relationship programs with customers, an appropriate balance between relationship costs with their customer and customer's profitability must be considered.

Cluster 3: this cluster form 10.5 percent of all customers, and contains 703 members. According to table 5, all variables of this model are higher than the data total average, and they are at the first position, and it shows a desirable cluster; so, they are loyal customers of a store. Therefore, for maintaining them specific programs should be considered, and more costs should be paid because more profitability of a store is obtained from this cluster.

4.6 Phase 6: Deployment

Step 7: Use of the presented model: At this stage, the model is re-examined and codified. In addition, the monitoring and maintenance plan is done after the completion of data mining. Finally, patterns become useful and usable knowledge. A customer (also known as a client, buyer, or purchaser) is the recipient of a good, service, product, or idea, obtained from a seller, vendor, or supplier for a monetary or other valuable consideration. Customers, clients, and buyers in construction may be any of the large and diverse range of private and public organizations, from government departments, large private corporations, individual customers, contractors, consultants, specialist suppliers (Preece, Chong, Golizadeh, & Rogers, 2015). Customers are the lifeblood of any organization. Without customers, a firm has no revenues, no profits, and therefore no market value (Gupta & Zeithaml, 2006, 718). With the increasing importance of the relationship with customers in modern commercial places, many companies are focused on

conversation, which is related to identification and customer profitability for earning customer satisfaction and loyalty. Clustering is a statistical technique that is used for minimizing different variations inside one cluster and maximizing variation between clusters according to distance or similarity criterion. In recent research, RFM model is used for customers clustering at a chain store, and its monetary volume variable (m) is corrected by the CPI index.

This corrected model has advantages over the main model; these advantages are contained: monetary changes at a period of time are identified, also according to clustering. Better quality of the reference model of m is done and more accurate identification of customers may be possible. So, chain stores, which are needy for analysis and identification of customers and for an information system that is able to save information about customer's transactions, can have qualified clustering on customers by applying this corrected model, and for each cluster, appropriate strategies of relationship with customers and marketing should be codified. It is suggested that in future researches complex wider data will be used so appropriate results can be obtained, and also it is offered that other algorithms of clustering will be used.

5. Conclusion and recommendations

This research identified the valuable customers for the shop, and it gives them a chance to choose goal customers and invest in them, but it doesn’t mean that the shop doesn’t pay attention to the other customers, and they limited their effort to satisfy their profitable customers but it means that they allocated a suitable budget for performing the relationship programs with customers to increase their loyalty and satisfaction. Using DM in CRM can be very useful in the following cases. It is suggested that in future researches complex wider data will be used so appropriate results can be obtained, and also it is offered that other algorithms of clustering will be used.

References

Aghimien, D., Aigbavboa, C., & Oke, A. (2019). A review of the application of data mining for sustainable construction in Nigeria. *Energy Procedia*, *158*, 3240–3245. https://doi.org/10.1016/j.egypro.2019.01.996

Alizadeh Zoeram, A., & Karimi Mazidi, A. R. (2018). New Approach for Customer Clustering by Integrating the LRFM Model and Fuzzy Inference System. *Iranian Journal of Management Studies*, *11*(2), 351–378.

Anshari, M., Almunawar, M. N., Lim, S. A., & Al-Mudimigh, A. (2018). Customer relationship management and big data enabled: Personalization & customization of services. *Applied Computing and Informatics*. https://doi.org/10.1016/j.aci.2018.05.004

Barson, R. J., Foster, G., Struck, T., Ratchev, S., Pawar, K., Weber, F., & Wunram, M. (2000). Inter-and intra-organisational barriers to sharing knowledge in the extended supply-chain. *Proceedings of the EBusiness and EWork*, 18–20.

Bose, I., & Chen, X. (2015). Detecting the migration of mobile service customers using fuzzy clustering. *Information & Management*, *52*(2), 227–238.

Burns, C., Sacks, G., & Gold, L. (2008). Longitudinal study of Consumer Price Index (CPI) trends in core and non-core foods in Australia. *Australian and New Zealand Journal of Public Health*, *32*(5), 450–453.

Buttle, F. (2009). *Customer Relationship Management: Concepts and Technologies*. Elsevier, Oxford.

Carrasco, R. A., Blasco, M. F., & Herrera-Viedma, E. (2015). A 2-tuple fuzzy linguistic RFM model and its implementation. *Procedia Computer Science*, *55*, 1340–1347. https://doi.org/10.1016/j.procs.2015.07.118

Casidy, R. (2012). Discovering consumer personality clusters in prestige sensitivity and fashion consciousness context. *Journal of International Consumer Marketing*, *24*(4), 291–299.

Chance, W. A. (1966). A note on the origins of index numbers. *The Review of Economics and Statistics*, 108–110.

Chang, H., & Tsay, S. (2004). Integrating od SOM and K-mean in data mining clustering: An empirical study of CRM and profitability evaluation. *Journal of Information Management*, *11*, 161–203.

Chemchem, A., & Drias, H. (2015). From data mining to knowledge mining: Application to intelligent agents. *Expert Systems with Applications*, *42*(3), 1436–1445.

Cheng, C. H., & Chen, Y. S. (2009). Classifying the segmentation of customer value via RFM model and RS theory. *Expert Systems with Applications*, *36*(3 PART 1), 4176–4184. https://doi.org/10.1016/j.eswa.2008.04.003

Christy, A. J., Umamakeswari, A., Priyatharsini, L., & Neyaa, A. (2018). RFM ranking – An effective approach to customer segmentation. *Journal of King Saud University - Computer and Information Sciences*. https://doi.org/10.1016/j.jksuci.2018.09.004

Dolca, I., & Nicolov, M. (2013). Analysis of Relationship between Net Wage and Consumer Price Index. *Procedia Economics and Finance*, *6*, 738–747.

Faezy razi, F., & Shadloo, N. (2017). A Hybrid Grey based Two Steps Clustering and Firefly Algorithm for Portfolio Selection. *Journal of Optimization in Industrial Engineering*, *10*(22), 49–59.

Garson, G. D. (2009). *Cluster analysis: statnotes*. Retrieved from http://faculty.chass.ncsu.edu/garson/pa765/statnote.htm

Ghalenooie, M. B., & Sarvestani, H. K. (2016). Evaluating Human Factors in Customer Relationship Management Case Study: Private Banks of Shiraz City. *Procedia Economics and Finance*, *36*, 363–373. https://doi.org/10.1016/S2212-5671(16)30048-X

Greenberg, P. (2010). *CRM at the Speed of Light: Social CRM Strategies, Tools, and Techniques*. McGraw-Hill.

Gupta, G. P. (2018). Improved Cuckoo Search-based Clustering Protocol for Wireless Sensor Networks. *Procedia Computer Science*, *125*, 234–240.

Gupta, S., & Zeithaml, V. (2006). Customer metrics and their impact on financial performance. *Marketing Science*, *25*(6), 718–739.

Haislip, J. Z., & Richardson, V. J. (2017). The effect of Customer Relationship Management systems on firm performance. *International Journal of Accounting Information Systems*, *27*, 16–29. https://doi.org/10.1016/j.accinf.2017.09.003

Harwati, A. M., & Karunia, A. (2017). Utilization of Social Media for Consumer Behavior Clustering using Text Mining Method. *Journal of Engineering and Applied Sciences*, *12*(3 SI), 6406–6411.
https://doi.org/10.3923/jeasci.2017.6406.6411

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

Imani, A., & Abbasi, M. (2017). Customers Clustering Based on RFM Model by Using Fuzzy C-means Algorithm (Case Study: Zahedan City Refah Chain Store). *Public Management Researches*, *10*(37), 251–276.
https://doi.org/10.22111/JMR.2017.3686

Kochetov, V. (2018). Overview of different approaches to solving problems of Data Mining. *Procedia Computer Science*, *123*(123), 234–239. https://doi.org/10.1016/j.procs.2018.01.036

Li, G., & Sun, L. (2018). Characterizing Heterogeneity in Drivers' Merging Maneuvers Using Two-Step Cluster Analysis. *Journal of Advanced Transportation*, *2018*, 1–15. https://doi.org/10.1155/2018/5604375

Li, Y., Huang, J., & Song, T. (2019). Examining business value of customer relationship management systems: IT usage and two-stage model perspectives. *Information & Management*, *56*(3), 392–402.

Merceron, A., & Yacef, K. (2004). Clustering students to help evaluate learning. *IFIP World Computer Congress, TC 3*, 31–42. Springer.

Mohammadian, M., & Makhani, I. (2016). RFM-Based customer segmentation as an elaborative analytical tool for enriching the creation of sales and trade marketing strategies. *International Academic Journal of Accounting and Financial Management*, *3*(6), 21–35.

Mohammadzadeh, M., Hoseini, Z. Z., & Derafshi, H. (2017). A data mining approach for modeling churn behavior via RFM model in specialized clinics Case study: A public sector hospital in Tehran. *Procedia Computer Science*, *120*, 23–30. https://doi.org/10.1016/j.procs.2017.11.206

Mozaffari A., Bonyadi A., A. F. (2010). Clustering and discovering the damage patterns of the third party insurance customers by using data mining techniques. *The 10th International Conference on Industrial Engineering*. Elsevier.

Ngai, E. W. T., Xiu, L., & Chau, D. C. K. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, *36*(2), 2592–2602.

Peker, S., Kocyigit, A., & Eren, P. E. (2017). LRFMP model for customer segmentation in the grocery retail industry: A case study. *Marketing Intelligence & Planning*, *35*(4), 544–559.

Preece, C., Chong, H. Y., Golizadeh, H., & Rogers, J. (2015). A review of customer relationship (CRM) implications: benefits and challenges in construction organizations. *International Journal of Civil Engineering*, *13*(3), 362–371.

Rajeh, S. M., Koudehi, F. A., Seyedhosseini, S. M., & Farazmand, R. (2014). A model for customer segmentation based on loyalty using data mining approach and fuzzy concept in Iranian Bank. *International Journal of Business and Behavioral Sciences*, *4*(9), 118–136.

Rémy, N. M., Martial, T. T., & Clémentin, T. D. (2018). The prediction of good physicians for prospective diagnosis using data mining. *Informatics in Medicine Unlocked*, *12*, 120–127. https://doi.org/10.1016/j.imu.2018.07.005

Safari, F., Safari, N., & Montazer, G. A. (2016). Customer lifetime value determination based on RFM model. *Marketing Intelligence & Planning*, *34*(4), 446–461.

Şchiopu, D. (2010). Applying TwoStep cluster analysis for identifying bank customers' profile. *Buletinul*, *62*, 66–75.

Soltani, Z., Zareie, B., Milani, F. S., & Navimipour, N. J. (2018). The impact of the customer relationship management on the organization performance. *The Journal of High Technology Management Research*, *29*(2), 237–246.

Song, M., Zhao, X., E, H., & Ou, Z. (2017). Statistics-based CRM approach via time series segmenting RFM on large scale data. *Knowledge-Based Systems*, *132*, 21–29. https://doi.org/10.1016/j.knosys.2017.05.027

Talón-Ballestero, P., González-Serrano, L., Soguero-Ruiz, C., Muñoz-Romero, S., & Rojo-Álvarez, J. L. (2018). Using big data from Customer Relationship Management information systems to determine the client profile in the hotel sector. *Tourism Management*, *68*, 187–197. https://doi.org/10.1016/j.tourman.2018.03.017

Tanaka, T., Hamaguchi, T., Saigo, T., & Tsuda, K. (2017). Classifying and Understanding Prospective Customers via Heterogeneity of Supermarket Stores. *Procedia Computer Science*, *112*, 956–964. https://doi.org/10.1016/j.procs.2017.08.133

Tekień, A., Gutkowska, K., Żakowska-Biemans, S., Jóźwik, A., & Krotki, M. (2018). Using cluster analysis and choice-based conjoint in research on consumers preferences towards animal origin food products. Theoretical review, results and recommendations. *Animal Science Papers and Reports*, *36*(2), 171–184.

Tripathi, S., Bhardwaj, A., & Poovammal, E. (2018). Approaches to Clustering in Customer Segmentation. In *International Journal of Engineering & Technology* (Vol. 7). https://doi.org/10.14419/ijet.v7i3.12.16505

Triznova, M., Maťova, H., Dvoracek, J., & Sadek, S. (2015). Customer Relationship Management Based on Employees and Corporate Culture. *Procedia Economics and Finance*, *26*, 953–959. https://doi.org/10.1016/s2212-5671(15)00914-4

Wang, Y., Assogba, K., Liu, Y., Ma, X., Xu, M., & Wang, Y. (2018). Two-echelon location-routing optimization with time windows based on customer clustering. *Expert Systems with Applications*, *104*, 244–260. https://doi.org/10.1016/j.eswa.2018.03.018

Wang, Y., Ma, X., Lao, Y., & Wang, Y. (2014). A fuzzy-based customer clustering approach with hierarchical structure for logistics network optimization. *Expert Systems with Applications*, *41*(2), 521–534.

Wu, H.-H., Chang, E.-C., & Lo, C.-F. (2009). Applying RFM model and K-means method in customer value analysis of an outfitter. In *Global Perspective for Competitive Enterprise, Economy and Ecology* (pp. 665–672). Springer.

Xiao, J., Wang, M., Tian, L., & Zhen, Z. (2018). The measurement of China's consumer market development based on CPI data. *Physica A: Statistical Mechanics and Its Applications*, *490*, 664–680. https://doi.org/10.1016/j.physa.2017.08.135

Zerbino, P., Aloini, D., Dulmin, R., & Mininno, V. (2018). Big Data-enabled customer relationship management: A holistic approach. *Information Processing & Management*, *54*(5), 818–846.

Zhang, T., Ramakrishnan, R., & Livny, M. (1996). BIRCH: an efficient data clustering method for very large databases. *ACM Sigmod Record*, *25*(2), 103–114. ACM.

Chou, Tung-Hsiang and Shu-Chen Chang. "The RFM Model Analysis for VIP Customer: A Case Study of Golf Clothing Brand." *IJKM* vol.18, no.1 2022: pp.1-18. http://doi.org/10.4018/IJKM.290025

Hu, Dingding, Zhou, Kaile, Li, Fangyi & Ma, Dawei. (2022). Electric vehicle user classification and value discovery based on charging big data. Energy,249,