Received 6 March 2021
Revised 20 April 2021
Accepted 7 May 2021

A multi-attribute data mining model for rule extraction and service operations benchmarking

Hannan Amoozad Mahdiraji
Leicester Castle Business School, De Montfort University, Leicester, UK
Madjid Tavana
Business Systems and Analytics Department, La Salle University, Philadelphia, Pennsylvania, USA and
Business Information Systems Department, University of Paderborn, Paderborn, Germany, and
Pouya Mahdiani and Ali Asghar Abbasi Kamardi
Faculty of Management, University of Tehran, Tehran, Iran

Abstract

Purpose – Customer differences and similarities play a crucial role in service operations, and service industries need to develop various strategies for different customer types. This study aims to understand the behavioral pattern of customers in the banking industry by proposing a hybrid data mining approach with rule extraction and service operation benchmarking.

Design/methodology/approach – The authors analyze customer data to identify the best customers using a modified recency, frequency and monetary (RFM) model and *K*-means clustering. The number of clusters is determined with a two-step *K*-means quality analysis based on the Silhouette, Davies–Bouldin and Calinski–Harabasz indices and the evaluation based on distance from average solution (EDAS). The best–worst method (BWM) and the total area based on orthogonal vectors (TAOV) are used next to sort the clusters. Finally, the associative rules and the Apriori algorithm are used to derive the customers' behavior patterns.

Findings – As a result of implementing the proposed approach in the financial service industry, customers were segmented and ranked into six clusters by analyzing 20,000 records. Furthermore, frequent customer financial behavior patterns were recognized based on demographic characteristics and financial transactions of customers. Thus, customer types were classified as highly loyal, loyal, high-interacting, low-interacting and missing customers. Eventually, appropriate strategies for interacting with each customer type were proposed.

Originality/value – The authors propose a novel hybrid multi-attribute data mining approach for rule extraction and the service operations benchmarking approach by combining data mining tools with a multilayer decision-making approach. The proposed hybrid approach has been implemented in a large-scale problem in the financial services industry.

Keywords Data mining, Rule extraction, *K*-means clustering, Evaluation based on distance from average solution, Total area based on orthogonal vectors, Best–worst method

Paper type Research paper

1. Introduction

Satisfied customers are the key to a successful business, and a deep understanding of consumer expectations is critical for long-term success. The financial services industry has realized the significance of managing customer relationships due to high levels of competition

The authors would like to thank the anonymous reviewers and the editor for their insightful comments and suggestions.

Dr. Madjid Tavana is grateful for the partial support he received from the Czech Science Foundation (GACR19-13946S) for this research.

Declaration of interest: The above authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Benchmarking: An International Journal
Vol. 29 No. 2, 2022
pp. 456-495

1463-5771
DOI 10.1108/BIJ-03-2021-0127

(Öztaysi *et al.*, 2011; Worthington and Welch, 2011). Loyalty in financial services is now more customer-centric, and customers expect to receive services that work for them (Leverin and Liljander, 2006; Reichstein and Salter, 2006). Customer segmentation enables companies to group customers and understand their needs. It also allows for individual and productive interaction with each group (Yao *et al.*, 2014). Research shows 1% growth in customer retention could improve the value of an organization by as much as 5% (Farajian and Mohammadi, 2010). Therefore, customer loyalty and retention is a valuable strategy that guarantees long-term value to organizations. This is more perceptible in financial services organizations where a large percentage of the business originates from a small percentage of customers (Cheng and Chen, 2009). Furthermore, the success of selling products to existing customers is usually much higher than selling to new customers. Numerous methods have been proposed for customer segmentation in the literature. Among these methods, clustering is the most common pre-institutional approach (Wedel and Kamakura, 2000). Moreover, the recency, frequency and monetary (RFM) model has also been used to understand and group customers according to the activities and buying behavior (Fader *et al.*, 2005; Newell, 1997). These methods have been widely used for customer analysis because of their simplicity, transparency and applicability.

Today, customers with different needs enter the markets, and various product or service providers are looking to attract them. Hence, service industries need to develop various marketing strategies for different customer types instead of a mass marketing strategy (Çınar *et al.*, 2020). Researchers have used various approaches like data mining (Çınar *et al.*, 2020), multi-criteria decision-making (MCDM) (Ghorabaee *et al.*, 2017a, b) and metaheuristic algorithms (Kuo *et al.*, 2020) for market segmentation. However, the quantity and the quality of the segments have not been studied thoroughly in the literature. Furthermore, the customers' behavior in each segment has not been analyzed systematically to ensure that the strategies conform to these behaviors.

The competition to attract new customers has intensified in the banking sector because some banks have merged to raise capital and increase their customer base. Lack of attention to new marketing principles such as customer development, new service delivery and target customer recognition in this competitive environment has resulted in customer loss in the financial services industry. Comprehensive customer service and support is the best strategy for customer loyalty and retention. This study proposes an integrated framework with the *K*-means clustering and the evaluation based on distance from average solution (EDAS) method to improve clustering quality and select the optimal number of customer clusters. Furthermore, we use the best-worst method (BWM) and the total area based on orthogonal vectors (TAOV) to rank the customer clusters. Finally, we use the Apriori algorithm to explore customer behavior patterns. By integrating these models with the customer lifetime value (CLV) model, we formulate strategies for each customer cluster and move from mass marketing to limited (individual) marketing for customer retention.

The remainder of this paper is organized as follows: Section 2 presents the theoretical foundations by discussing the significance of big data applications, customer segmentation, the RFM method, *K*-means clustering, the BWM, EDAS, TAOV and association rules. In Section 3, we introduce the proposed integrated framework, and in Section 4, we present our analysis and results. Finally, in Section 5, we present our practical implications, and in Section 6, we present our conclusions.

2. Theoretical foundations

2.1 Significance of big data application

Nowadays, the impressive growth of data may be reached from any source, for example, sensors, shopping transactions and even social media networks. The speed of data advancement has indeed surpassed Moore's law (Chen and Zhang, 2014). Each day in 2011,

more than 2.5 centillion data have been produced according to International Business Machines Corporation (IBM) reports (Hilbert and López, 2011). Figure 1 elaborates on the results of the global data forecast provided by the International Data Corporation (IDC). There is no doubt that the era of big data has arrived (Wang *et al.*, 2016). In addition to high volume, big data is related to structural complexity, the complication of data acquisition and data management (Casado and Younas, 2015).

Gutner indicates that big data will become one of the top ten technologies of the next five years, in a report in 2012. Big data creates huge value. These values are created as a chain through the processes of data discovery, integration and exploitation (Miller and Mork, 2013). The McKinsey Institute has reported that over 50% of the 560 surveyed companies stressed that big data could help the selection of appropriate strategies and customer services (Manyika *et al.*, 2011). This big data can support smart organizations' decisions. However, they also need to support themselves. Figure 2 demonstrates the big data value chain.

As shown in Figure 2, big data–driven decision-making involves data acquisition, data preparation, data analysis, data visualization and informed decision-making. There are numerous definitions of big data formation (Ekbia *et al.*, 2015). The product-oriented aspect emphasized data characteristics such as their size, speeds and structures (Gobble, 2013). The process-oriented viewpoint highlights the characteristics of the processes involved in the storage, management, collection, search and analysis of big data (Jacobs, 2009; Kraska, 2013). The cognition-based approach focuses on the challenges posed by big data due to their cognitive capacities and limitations (Manyika *et al.*, 2011). Finally, the social movement perspective draws attention to the gap between view and reality, particularly the social, economic, cultural and political movements that show the existence of big data (Ekbia *et al.*, 2015). There are usually three main activities related to the high-focused data, including the acquisition, modification and analysis of big data (Tansley and Tolle, 2009). Nevertheless, the purpose of big data processing is to exploit the knowledge extracted from the data to protect smart decision-making. The process of knowledge discovery from this data includes the steps of data collection, data preprocessing (cleaning and integration), data conversion, data mining and interpretation, and data evaluation [e.g. visualization (Bhambri, 2011; Newton, 2004)]. Figure 3 displays data mining tools.

As shown in Figure 3, there are two different types of data mining tools: descriptive and predictive. The descriptive tools (i.e. associative rules, summarization, pattern recognition and clustering) are used to scrutinize the data, and the predictive tools (i.e. regression, time series, forecasting and classification) are used to predict a relevant behavior or pattern in the data. More specifically, data mining is defined as an advanced information search that includes a statistical algorithm to discover patterns and relationships of data (Topi and Tucker, 2014). Data mining tools capture data and create a model of reality as a model. The resulting model describes the motifs and relationships of the data. Data mining tools pictured

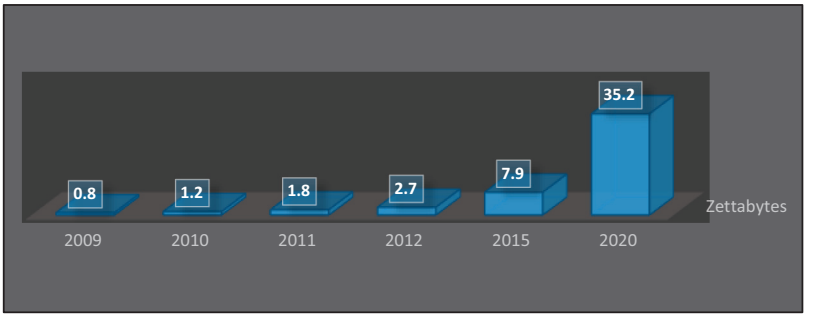


Figure 1.
The global data volume forecast

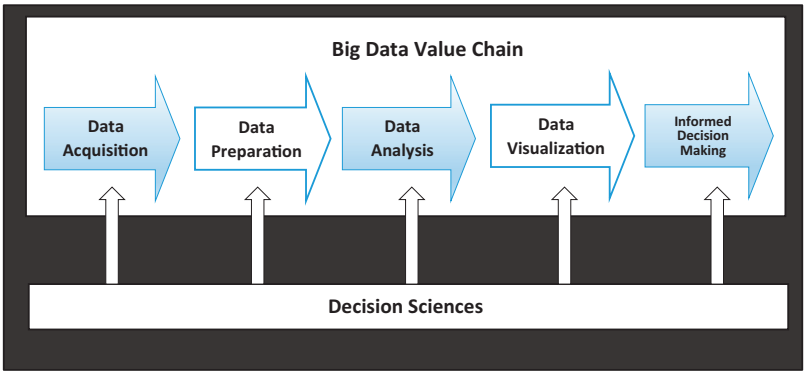


Figure 2.
Big data value chain

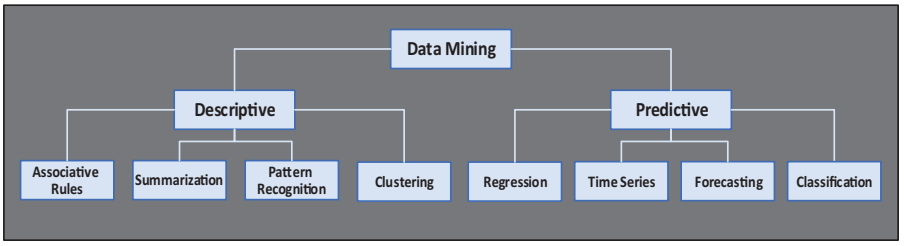


Figure 3.
Data mining tools

above are used in various businesses and industries. Banks can employ the knowledge discovery process by data mining methods for plenty of their operations, including card marketing, pricing and profitability from card owners, fraud discovery, predictive life cycle management and customer segmentation (Mahdiraji *et al.*, 2019). Among these various applications of data mining in banking, customer segmentation is presented further.

2.2 Customer segmentation

Customers have divergent needs, behaviors and preferences. Equal treatment and service to these heterogeneous customers are challenging for companies (Peker *et al.*, 2017). Customer segmentation was first introduced by Smith in 1956 and has emerged to address this problem (Smith, 1956). Subsequently, it was supported by abundant companies in several fields. Customer segmentation is dividing the entire customer pool into smaller segments, each containing customers that have similar requirements and characteristics. In customer segmentation, the variables are generally separated into two categories. General variables include customer demographics (e.g. gender, age, income, education level) and lifestyle (e.g. urban, rural). Besides, specific product-based variables contain customer purchase behavior (e.g. number of purchases, uses, costs) and intentions. Either way, it is easier to stress and work with general variables. However, specific product and service-centered variables are more highlighted for understanding customer behavior and provide more opportunity to engage in differentiating customers to deal with them (Tsai and Chiu, 2004). Wide use of customer segmentation has been noted in recent years. As an illustration, the segmentation cashback website customers in the field of e-commerce was presented (Ballestar *et al.*, 2018). Furthermore, air cargo customers were segmented by an intelligent model to maintain customers and increase profit margins (Yin *et al.*, 2019). Moreover, customer loyalty in the

field of local brand fashion was discussed by concentrating on customer segmentation (Dachyar *et al.*, 2019). Furthermore, market segmentation in willingness to order private label brands was illustrated from the viewpoint of e-grocery shoppers (Jagani *et al.*, 2020).

In this context, features of the RFM model and the derivatives derived from this model are recognized as highly applicable and effective. Hence, it has been applied in bountiful studies to identify customer treatments. Considering the customers' needs and paying attention to them can help companies maintain long-term relationships with their customers (Dibb, 1998). Moreover, companies can improve their revenues by attracting and retaining valuable customers at the lowest cost (Safari *et al.*, 2016). Due to the significance of market segmentation, numerous techniques have been applied to propose an appropriate model to perform it.

2.3 Recency, frequency and monetary model for predicting customer behavior

The RFM model was first introduced by Hughes in 1996 to analyze and predict customer behavior. The basic RFM model has three main functions:

(1) *Recency (R).* The recency or recent transactions refers to the time from the last purchase (day or month). It provides information about the potential presence of the customer to repurchase. The probability of a repurchase or represence of the customer will be higher for nearer recency.

(2) *Frequency (F).* Frequency is the number of times that a customer has been present. In other words, it points to the existence of a customer over a particular period, which indicates customer loyalty. Higher purchase frequency is a sign of higher customer loyalty.

(3) *Monetary value (M).* Monetary value consists of the total amount of money spent. By way of explanation, it declares the average amount of money spent over a precise period that assesses the customer's share of a company's revenue. The larger this value, the greater the customer's share of the company's revenue.

Some extensions and similar versions of the RFM model are as follows:

(1) *RFMTC* for adding two other variables, including time of first purchase and churn probability to recency, frequency and monetary.

(2) *Timely RFM*, or TRFM, refers to the periodicity of the product.

(3) *RFD* signifies recency, frequency and duration, which considers the length of time for which a website has been visited (Yan and Chen, 2011).

(4) *RML*, includes recency, monetary valueand loyalty; it adapted the RFM model to the annual transaction environment (Dursun and Caber, 2016).

(5) *RFR* consists of recency, frequency and reach, which are intended for social elements such as last post (post) and repetition (number of posts) or access (i.e. reach a network of friends) (Birant, 2011).

(6) *FRAT* contains frequency, recency, amount and product types that perform the sorting task according to the types of products purchased (Woo *et al.*, 2005).

The previous research shows the widespread use of the *K*-means clustering algorithm by the side of RFM models. An adapted RFM model to estimate passenger value in Taiwan was previously used (Wong *et al.*, 2006). Furthermore, the RFM model to predict target market value in travel clubs was designed (Lumsden *et al.*, 2008). Besides, customer value found on

the weighted RFM model to decide on the customer relationship management (CRM) system in the banking industry was calculated (Khajvand *et al.*, 2011). Alongside the RFM model, the *K*-means algorithm is applied in the field of classification and segmentation (Cheng and Chen, 2009). Moreover, self-organizing maps (SOM) and *K*-means with the RFM model in the hair salon area of Taiwan were employed (Wei *et al.*, 2010). Moreover, the customer's purchase treatments by considering market segmentation are analyzed recently (Anitha and Patil, 2019).

2.4 K-*means clustering*

Clustering is a significant tool that has been popularly employed in customer segmentation (Chiu and Tavella, 2008; Sarstedt and Mooi, 2014). The purpose of clustering is to group a set of objects that have the most similarity of properties among each other (Jain *et al.*, 1999). Figure 4 has pictured the types of clustering methods (see Figure 5).

As shown in Figure 4, clustering methods are classified into hierarchical and partitioned methods. Hierarchical methods are used when the relationship among the clusters is dependent, and partitioned methods are used when the relationship among the clusters is independent. Furthermore, hierarchical methods are divided into agglomerative and divisive models, and partitioned methods are divided into distance-based, model-based and density-based models. The *K*-means algorithm (MacQueen, 1967) is one of the most common techniques (Jain, 2010). This algorithm can be run easily and rapidly (Cheung, 2003; Davidson, 2002). This method requires specifying the number of clusters (*K*). The steps to implement the *K*-means algorithm are as follows:

(1) The *K*-points are selected as the points of the centroids of the clusters.

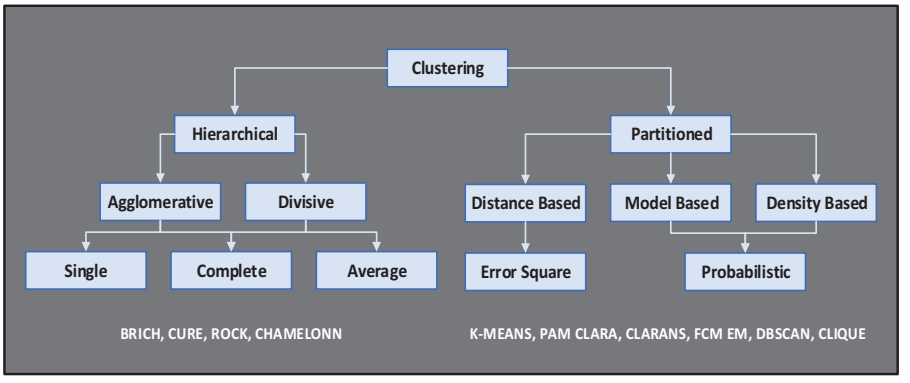
(2) Each data sample is assigned to the cluster whose center has the least distance to that data. These distances can be computed by Euclidean as follows.

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} ||x_i - c_j||^2 \tag{1}$$

Notably, J demonstrates the distance between the i_{th} data sample (x_i) and the center of the j_{th} cluster (c_j). Moreover, n is the number of the data sample, and k is the number of clusters.

(3) After all, data belongs to one cluster; a new point is calculated for each cluster as the center.

(4) Steps 2 and 3 are repeated until no change in the center of the clusters is achieved.



Source(s): Saxena *et al.* (2017)

Figure 4.
Clustering methods

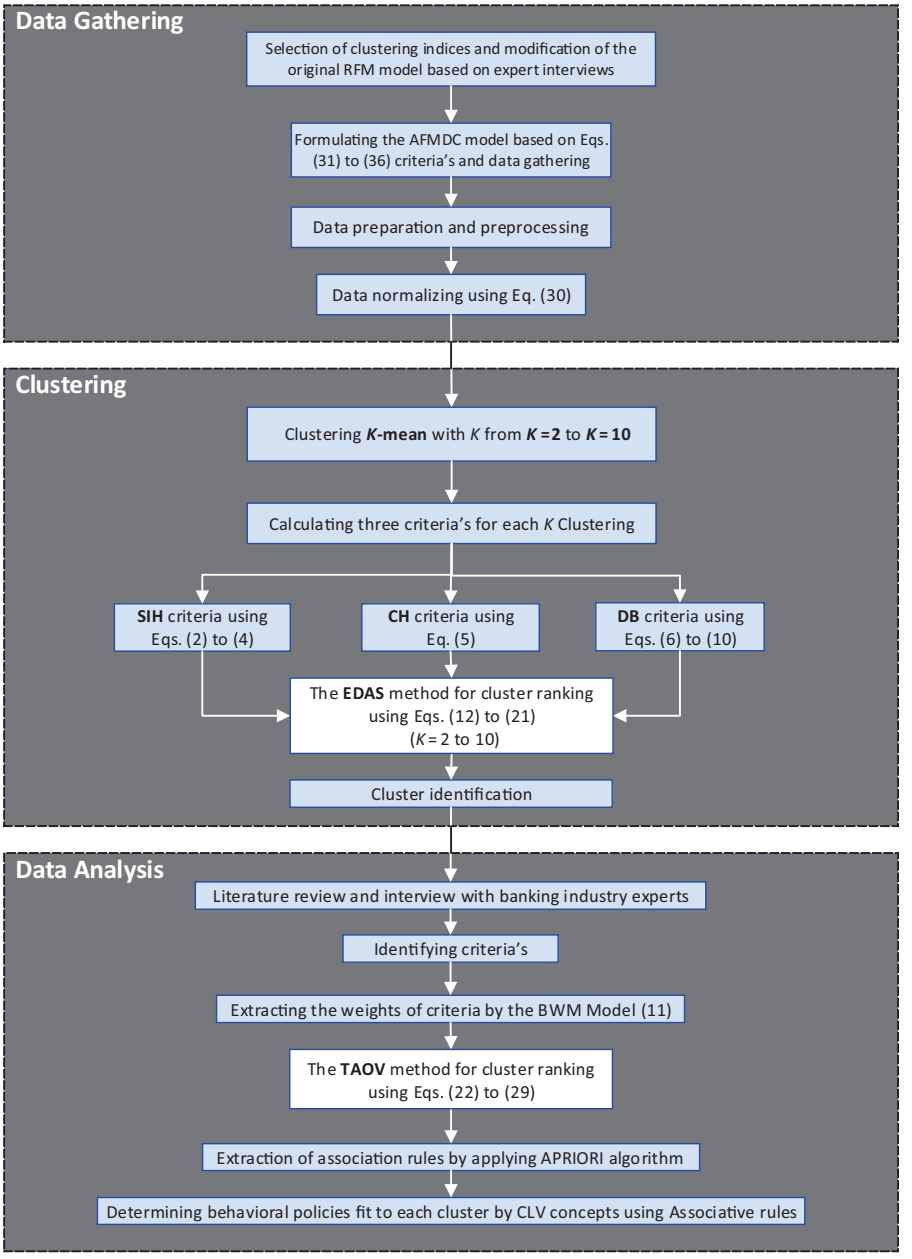


Figure 5.
The proposed framework

K-means clustering has been applied in numerous fields with Convex Hux to predict future disasters (Basak *et al.*, 2019; Gupta *et al.*, 2019). Furthermore, *K*-means is employed to detect a brain tumor (Khan *et al.*, 2019). Notably, customer research and analysis are some highlighted fields in which *K*- means have been used. There are miscellaneous criteria for evaluating

clusters. Three of these criteria which are frequently used are silhouette (SIH), Davies–Bouldin (DB) and Calinski–Harabasz (CH).

Silhouette: It is one of the methods to evaluate clustering (Rousseeuw, 1987). This criterion depends both on the cohesion of the clusters and the degree of their distinction. The value of SIH for each point measures its belonging to the cluster in comparison with the belonging to the adjacent cluster. The focus of the SIH criterion relies on the quality of the performed clustering. This criterion determines what the data distribution is in the clusters. The higher the SIH, the better the clustering quality. The following two concepts are measured in calculating the SIH:

(1) *Mean distance of a point to other points within a cluster:* Suppose x_i belongs to the C_j cluster. The mean distance of this point to other points within the cluster will be measured by Eq. (2).

$$a(i) = \frac{1}{n_j} \sum_{i=1}^{n_j} d(x_i, x_l) \tag{2}$$

Note that n_j is the size of cluster j. Besides $d(x_i, x_l)$ demonstrate the distance between data sample $x_l \in C_j$ and other data samples in cluster j (x_i). Indeed $a(i)$ elaborates the belonging value of the x_i to its cluster, which is more for the lower values. This distance can be measured by divergent functions, for example, Manhattan and Euclidean functions.

(2) *Minimum of a mean distance of a point to other clusters:* Suppose that x_i is a point belonging to cluster C_j. The minimum of a mean distance of this point to cluster C_k is computed by Eq. (3).

$$b(i) = \min_{1 \le l \le k} \frac{1}{n_l} \sum_{y_m \in c_k} d(x_i, y_m) \tag{3}$$

It is notable that $d(x_i, y_m)$ is the distance between data sample x_i and y_m which are the points belonging to C_k. Moreover, n_l is the number of measured distances. A cluster that has the lowest mean distance to the point x_i is referred to as an adjacent cluster to this point. Thus, the value of the SIH criterion $s(i)$ for point x_i is calculated by Eq. (4).

$$s(i) = \frac{b(i) - a(i)}{\max(b(i), a(i))} \tag{4}$$

where $a(i)$ is the mean distance of this point to other points within the cluster which is measured by Eq. (2), and $b(i)$ is the minimum of a mean distance of this point to other clusters that are calculated by Eq. (3).

Calinski–Harabasz Index: For a set of data E by the size of N, which is grouped into K clusters, this index is called the variance ratio criterion (VRC) and is computed by Eq. (5).

$$\mathrm{VRC}_k = \frac{\mathrm{SS}_B}{\mathrm{SS}_W} \times \frac{(N-K)}{(K-1)} \tag{5}$$

Note that SS_B is the sum of variance between clusters obtained by Eq. (6), and SS_W is the sum of variance within clusters measured by Eq. (7).

$$\mathrm{SS}_B = \sum_{j=1}^{k} n_j (c_j - c_E)(c_j - c_E)^T \tag{6}$$

$$SS_W = \sum_{j=1}^{k} \sum_{x_i \in c_j} (x_i - c_j)(x_i - c_j)^T \tag{7}$$

In Eq. (6), n_j is the size of cluster j, c_j is the centroid of cluster j and c_E is the centroid of E. Also, in Eq. (7), x_i represents the members of cluster j. The best number of clusters is determined by the largest VRC value (Calinski and Harabasz, 1974). It is illustrious that selecting the optimal number of clusters is one of the most remarkable steps in implementing clustering. MCDM is a beneficial tool to find the optimal number of clusters.

Davies–Bouldin Index: This index was introduced by Davies and Bouldin, two scientists in the field of electricity in 1979 (Davies and Bouldin, 1979). This index is not dependent on the number of clusters or the clustering algorithm. The two criteria introduced next are used to calculate this index.

(1) *The measure of scattering within a cluster:* Suppose that S_i is the measure of scattering corresponding to the cluster C_i, and d is also a distance function. Then the scattering rate for this cluster will be calculated by Eq. (8) in case (r) denotes the order.

$$S_i = \left[\frac{1}{|c_i|} \sum_{x \in c_i} d^r(x, c_i) \right]^{\frac{1}{r}} \tag{8}$$

Note that C_i is the centroid of cluster i. This relation is, in fact, similar to Minkowski's distance of the points of each cluster from its centroids.

(2) *Cluster separation:* The separation between the two clusters is also measured by the distance between their centroids. If V_i and V_j are the centroids of the clusters i and j and $d(v_i, v_j)^t$ is the distance between these two centroids considering the order of t, the distance between these two clusters is shown by D_{ij} and is obtained by Eq. (9) for a specific order of t.

$$D_{ij} = \left[\sum d(v_i, v_j)^t \right]^{\frac{1}{t}} \tag{9}$$

Considering S_i and S_j as the scatter of the cluster i and j obtained by Eq. (8), and D_{ij} for separation measured by Eq. (9), R_{ij} can be calculated by Eq. (10), which demonstrates how good the clustering scheme is.

$$R_{ij} = \frac{S_i + S_j}{D_{ij}} \tag{10}$$

The maximum distance of each cluster relative to the other clusters is computed by Eq. (11) to attain the DB Index for a clustering method.

$$R_i = \max R_{ij} \tag{11}$$

Then the mean of the maximum distances will be calculated for all clusters by Eq. (12), which indicates the DB Index.

$$V_{\mathrm{DB}} = \frac{\sum_i^k R_i}{k} \tag{12}$$

It is noteworthy that k is the number of clusters, and the lower the DB Index is, the better is the clustering. However, in this research, we proposed a decision-making approach to identify the optimal number of clusters. Stewart (1992) proposes that an MCDM method aims to assist

the decision-maker in discovering the solution to the problem. In the coming sections, some of the MCDM techniques used in this research are described, including BWM, EDAS and TAOV.

2.5 Best–worst method

The BWM was developed by Rezaei (2015). Many MCDM methods are available to calculate the weight and importance of criteria, such as the analytical hierarchical process (AHP), simultaneous evaluation of criteria and alternatives (SECA), etc. However, most of these methods are based on heuristic calculations and usually result in nearly optimal values. In this regard, the BWM is one of the newest and most effective MCDM techniques used to extract the weights of the criteria through a nonlinear mathematical model with global optimal values. The following steps are required to perform this technique:

(1) A set of decision criteria are elicited ($\{C_1, C_2 \ldots, C_n\}$).

(2) The best and the worst criterion are determined; the best can be the most desirable or the most crucial.

(3) Pair comparisons between the best criterion and the other criteria are made ($A_B = \{a_{b1}, a_{b2}, \ldots, a_{bm}\}$).

(4) Pair comparisons between the other criteria and the worst criterion are made ($A_w = \{a_{1w}, a_{2w}. \ldots, a_{nw}\}$).

(5) The optimal weights are extracted by solving the model of (13) ($\{W_1, W_2. \ldots, W_n\}$).

$$\begin{aligned} &\min \xi \\ &\text{st}: \\ &\left|\frac{W_B}{W_j} - A_{bj}\right| \le \xi; \qquad \text{for all } j \\ &\left|A_{jw} - \frac{W_j}{W_W}\right| \le \xi; \qquad \text{for all } j \\ &\sum W_j = 1 \\ &W_j \ge 0 \end{aligned} \tag{13}$$

Note that W_B and W_W represent the weight of the best and the worst criterion, and W_j represents the weight of the jth criteria. Moreover, the objective function (ξ) aims to minimize the difference between the W_j value resulted from the model and the A_{bj} and A_{jw} values provided by the experts. Besides, the compatibility rate of comparisons or consistency ratio (CR) is computed by $\text{CR} = \frac{\xi^*}{\text{CI}}$ to check the validity of the results derived from the experts. CR values less than or equal to 0.1 are acceptable. Note that *CI* is the Compatibility Index, which is determined based on the preference of the best criterion over the worst criterion (A_{BW}). The *CI* values are mentioned in Table 1.

It is preferred that the value of the CR for each expert result is less than 0.5. Higher values are not recommended, and the expert should fill the questionnaire again or replace it with another possible expert. Numerous applications of the BWM have been investigated in numerous fields. Some applied the BWM to evaluate key factors of sustainable architecture (Amoozad Mahdiraji *et al.*, 2018) or locate a hotel by a viewpoint of sustainability (Zolfani *et al.*, 2019). Recently, the combination of the BWM and combinative distance-based

assessment (CODAS) under the condition of interval-valued multi-granular 2-tuple linguistic was used for site selection in construction projects (Maghsoodi *et al.*, 2020).

2.6 Evaluation based on distance from an average solution technique

EDAS is an MCDM method to prioritize the alternatives. A wide range of methods are available in the MCDM to rank alternatives based on specific criteria. As in this research SIH, CH and DB criteria are used to evaluate the different number of clusters for customers based on the *K*-mean method, the distance between average values of these three criteria is critical. Hence, as in the EDAS method, two metrics are employed for alternative assessment containing positive distance from average (PDA) and negative distance from average (NDA). The executive steps are as follows (Ghorabaee *et al.*, 2017a, b):

(1) The most highlighted criteria that define alternatives are selected.

(2) The decision matrix is constructed. X_{ij} symbolizes the performance of the value of the i_{th} alternative over the j_{th} criterion (DM = $[X_{ij}]$).

(3) Average solutions are obtained by Eq. (14) (AV = $[\mathrm{AV}_j]$).

$$\mathrm{AV}_j = \frac{\sum_{i=1}^{n} x_{ij}}{n} \tag{14}$$

(4) PDA and NDA are measured by Eqs. (15) and (16) for beneficial criteria and Eqs. (17) and (18) for cost criteria.

$$\mathrm{PDA}_{ij} = \frac{\max(0, (x_{ij} - \mathrm{AV}_j))}{\mathrm{AV}_j} \tag{15}$$

$$\mathrm{NDA}_{ij} = \frac{\max(0, (\mathrm{AV}_j - x_{ij}))}{\mathrm{AV}_j} \tag{16}$$

$$\mathrm{PDA}_{ij} = \frac{\max(0, (\mathrm{AV}_j - x_{ij}))}{\mathrm{AV}_j} \tag{17}$$

$$\mathrm{NDA}_{ij} = \frac{\max(0, (x_{ij} - \mathrm{AV}_j))}{\mathrm{AV}_j} \tag{18}$$

(5) Find the weighted sum of PDA (sum of positive [SP]) and the weighted sum of NDA (sum of negative [SP]) for all alternatives by Eqs. (19) and (20). Note that W_j is the weight of the j_{th} criterion.

$$\mathrm{SP}_i = \sum_{j=1}^{m} w_j \times \mathrm{PDA}_{ij} \tag{19}$$

$$\mathrm{SN}_i = \sum_{j=1}^{m} w_j \times \mathrm{NDA}_{ij} \tag{20}$$

Table 1.
Consistency Index found on the preference of the best criterion over the worst criterion

A_{BW}	1	2	3	4	5	6	7	8	9
CI	0.00	0.44	1.00	1.63	2.30	3.00	3.73	4.47	5.23

(6) SP and SN values are normalized for all alternatives by Eqs. (21) and (22).

$$\mathrm{NSP}_i = \frac{\mathrm{SP}_i}{\max_i(\mathrm{SP}_i)} \tag{21}$$

$$\mathrm{NSN}_i = 1 - \frac{\mathrm{SN}_i}{\max_i(\mathrm{SN}_i)} \tag{22}$$

(7) The evaluation score for all alternatives (AS) is measured by Eq. (23) when $0 \leq \mathrm{AS} \leq 1$.

$$\mathrm{AS}_i = \frac{1}{2}(\mathrm{NSP}_i + \mathrm{NSN}_i) \tag{23}$$

(8) Criterion ratings are based on AS devaluation. The criteria with the highest values are the most valuable ones.

It is mentioned that NDA is zero for alternatives with positive PDA, and PDA is zero for alternatives with positive PDA. Abundant studies have been performed by the EDAS technique. Some scholars applied EDAS for green supplier selection (He *et al.*, 2019; Zhang *et al.*, 2019). This method is recently employed by the BWM and wavelet neural networks for cloud service selection (Gireesha *et al.*, 2020).

2.7 Total area based on the orthogonal vector method

The TAOV is an MCDM technique to rank alternatives (Hajiagha *et al.*, 2018). Scoring (e.g. AHP), compromising (e.g. EDAS) and outranking methods are available to rank different alternatives in the MCDM era. In this regard, the TAOV algorithm is used in three phases: initialization, segmentation and comparison. This method benefits from principal component analysis (PCA) to create the matrix of the distance between each pair of alternatives. The implementation steps in each phase are as follows:

(1) The decision alternatives are identified [$(A_1, A_2 \ldots, A_n)$].

(2) The decision criteria are recognized [$(C_1, C_2 \ldots C_n)$].

(3) decision matrix is created ($X = [X_{ij}]$).

(4) The weight vectors are found [$w = (w_1, w_2 \ldots w_n)$].

(5) decision matrix is normalized for cost–benefit criteria (C represent cost and B elaborates benefit criteria) by Eqs. (24) and (25) ($R = [r_{ij}]$),

$$r_{ij} = \frac{x_{ij}}{\max_i(x_{ij})} \quad ,j \in B \tag{24}$$

$$r_{ij} = \frac{\min_i(x_{ij})}{x_{ij}} \quad ,j \in C \tag{25}$$

(6) The weighted normalized matrix is calculated by Eq. (26) ($\mathrm{WN} = [\tilde{r}_{ij}]$),

$$\tilde{r}_{ij} = w_j \cdot r_{ij} \tag{26}$$

(7) The normalized weighted matrix is converted to the Y-equivalent matrix of (27) regarding the calculation of Eqs. (27) to (29) using the principal component analysis.

Moreover, the distance between every two elements is computed by Eq. (29).

$$Y^T = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \mathrm{AY}^t = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \cdot \begin{bmatrix} \overline{r_1} \\ \vdots \\ \overline{r_n} \end{bmatrix} \tag{27}$$

$$Y = \begin{bmatrix} \begin{bmatrix} y_{11} & \cdots & y_{1n} \\ \vdots & \ddots & \vdots \\ y_{m1} & \cdots & y_{mn} \end{bmatrix} \end{bmatrix} \tag{28}$$

$$d^i_{k,l} = \sqrt{y^2_{ik} + y^2_{il}} \tag{29}$$

(8) The attractiveness of each alternative is calculated by Eq. (30),

$$\mathrm{TA}_i = \sum_{j=1}^{n-1} d^i_{j,j+1} \tag{30}$$

(9) The attractiveness of alternatives is computed by applying the normalized total area by Eq. (31).

$$\mathrm{NTA}_i = \frac{\mathrm{TA}_i}{\sum_{k=1}^{m} \mathrm{TA}_k} \tag{31}$$

2.8 Association rules

Exploring association rules is one of the remarkable data mining techniques, perhaps the most common ones for finding local patterns of nonsupervised learning systems. These techniques can be very useful in predicting the behavior of customers. Association rules allow characterized conditional terms. An association rule consists of two sets of items:

(1) The antecedent or left-hand side (LHS),

(2) Consequent or right-hand side (RHS) which is combined with repeat-based statistics. It interestingly illustrates the relationship between support and confidence.

The Apriori algorithm is the most highlighted and classical algorithm for the discovery of repeated item sets (Tiwari *et al.*, 2010). The Apriori is used to find the entire set of data items in the provided database. Table 2 illustrates the previous research in the field of market segmentation, customer clustering, customer analysis and related topics.

Although numerous models have been proposed to segment the customers, these models have to be extended considering the type of industry, customers, accessible data and the aim of the segmentation. Inappropriate segments would lead to the loss of marketing investment. The quality of the segmentation is remarkably important and should be evaluated. An applicable model considering various characteristics of the market can extract more valuable information and knowledge from big data (Hu *et al.*, 2020). Moreover, the techniques that are applied to perform the segmentation play a meaningful role in improving the quality of analysis (Kimiagari *et al.*, 2021; Parikh and Abdelfattah, 2020). Alongside the quality, the quantity of the segments is an important indicator of assessing the segmentations (Munusamy and Murugesan, 2020). Furthermore, to evaluate the quality and quantity of the segments, analyzing the behavior of the customers in each segment is also critical (Reutterer *et al.*, 2020). Understanding their action will lead to developing more effective strategies for

Table 2. Previous research studies

(Researcher/ Year)	Objective	Methods													
		Bayes	RFM	*K*-mean/ medoids	CLV	CVM	FCM	MCA	Regression	Decision-making/ tree	Apriori	*C*-mean	SOM	FIS	WEKA/ WARD
Song *et al.* (2016)	Validating hidden customers applying time Series, segmentation found on the RFM model in big data, MCA and RFM integration		✓	✓				✓							
Estrella-Ramón *et al.* (2017)	Providing a CLV model, evaluating and classifying bank customers with individual measures	✓			✓	✓			✓						
Sivasankar and Vijaya (2017)	Customer segmentation, comparison of clustering methods In combination with decision tree – integration of supervisory and nonsupervisory methods			✓			✓			✓					

(*continued*)

Table 2.

(Researcher/ Year)	Objective	Methods													
		Bayes	RFM	*K*-mean/ medoids	CLV	CVM	FCM	MCA	Regression	Decision-making/ tree	Apriori	*C*-mean	SOM	FIS	WEKA/ WARD
Chiang (2017)	Taiwan air travel market segmentation, finding more valuable markets, extracting the right rules, improving CRM		✓								✓				
Peker *et al.* (2017)	Identifying distinct customer segments and clusters		✓	✓											
Öner and Öztayşi (2017)	Investigation of data analytics, understanding the similarities of customer shopping places, clustering places and customers						✓					✓			
Qadadeh and Abdallah (2018)	Investigation of data analytics algorithms especially *K*-means, SOM with a database (TIC)			✓									✓		

(*continued*)

Table 2.

(Researcher/ Year)	Objective	Methods													
		Bayes	RFM	*K*-mean/ medoids	CLV	CVM	FCM	MCA	Regression	Decision-making/ tree	Apriori	*C*-mean	SOM	FIS	WEKA/ WARD
Li *et al.* (2018)	Clustering customers, cluster analysis, formulating marketing strategies		✓	✓											
Alizadeh Zoeram and Karimi Mazidi (2018)	Offering a systematic approach to analyze customer behavior, improving the performance of the customer management system		✓		✓		✓							✓	
Wilson *et al.* (2018)	Discovery of association rules, clustering customers' unstable treatment (TBS), helping the design of smart campaigns			✓							✓				✓

(*continued*)

Table 2.

(Researcher/ Year)	Objective	Methods													
		Bayes	RFM	*K*-mean/ medoids	CLV	CVM	FCM	MCA	Regression	Decision-making/ tree	Apriori	*C*-mean	SOM	FIS	WEKA/ WARD
Zhang *et al.* (2018)	Offering an improved model of genetic algorithm and *K*-means clustering, extracting customer characteristics, comparing the proposed model to three other models		✓	✓							✓		✓		
Saeedi and Albadvi (2018)	Building a customer valuation model, considering financial value, structural value and influencing value, clustering customers into cohesive groups			✓	✓										
Aryuni and Miranda (2018)	Comparing the performance of *K*-means with *K*-medoids algorithms		✓	✓											

(*continued*)

Table 2.

(Researcher/ Year)	Objective	Methods													
		Bayes	RFM	*K*-mean/ medoids	CLV	CVM	FCM	MCA	Regression	Decision-making/ tree	Apriori	*C*-mean	SOM	FIS	WEKA/ WARD
De Caigny *et al.* (2018)	Improving the performance of decision tree, improving the performance of logistic regression and comparison of a novel model in RF and LMT								✓	✓					
Phan *et al.* (2019)	Discovery of customers' financial attitudes and behaviors in Switzerland and Vietnam			✓											✓
Maji *et al.* (2019)	Developing the plan of discovery and ELT of data, facilitating the estimation of the percentage of bank cardholders			✓											

(*continued*)

Table 2.

(Researcher/ Year)	Objective	Methods Bayes	RFM	*K*-mean/ medoids	CLV	CVM	FCM	MCA	Regression	Decision-making/ tree	Apriori	*C*-mean	SOM	FIS	WEKA/ WARD
Mahdiraji *et al.* (2019)	Proposing a model for Iranian banks to analyze and distinguish customers' needs for service suggestions		✓	✓						✓					
Motlagh *et al.* (2019)	Introducing a strategy to ease the limitations by converting any types of load time series into map models that could be readily clustered			✓											✓
Hu *et al.* (2020)	Presenting an RFMT customer classification model based on customer behavior		✓	✓											

(*continued*)

Table 2.

(Researcher/ Year)	Objective	Methods													
		Bayes	RFM	*K*-mean/ medoids	CLV	CVM	FCM	MCA	Regression	Decision-making/ tree	Apriori	*C*-mean	SOM	FIS	WEKA/ WARD
Parikh and Abdelfattah (2020)	Studying the performance of the RFM model and clustering in online transactions to provide strategies for customer purchasing behaviors		✓	✓											
Vohra *et al.* (2020)	Employing 2010 retail data to generate meaningful business intelligence		✓	✓									✓		
Rahmadianti *et al.* (2020)	Providing company insight to assess their customers and improve marketing strategies		✓	✓	✓					✓					

(*continued*)

Table 2.

(Researcher/ Year)	Objective	Methods													
		Bayes	RFM	*K*-mean/ medoids	CLV	CVM	FCM	MCA	Regression	Decision-making/ tree	Apriori	*C*-mean	SOM	FIS	WEKA/ WARD
Zhang *et al.* (2021)	Recognizing behavior observation and introducing an evaluation model from the perspective of market segmentation		✓		✓					✓					
De Marco *et al.* (2021)	Elaborating that applying cognitive analytics management methodology is a valid tool to describe new technology implementations for businesses		✓		✓										

Note(s): CLV, Customer Lifetime Value; CVM, Customer Value Management; FCM, Fuzzy Cognitive Mapping; MCA, Multiple Correspondence Analysis; RFM, Recency, Frequency, Monetary; SOM, Self-Organizing Map; FIS, Fuzzy Inference System; WEKA, Waikato Environment for Knowledge Analysis (Software); WARD, a method named by Ward, J. H., Jr.

each group. This research attempts to advance the market segmentation by developing the RFM model for predicting customer behavior. In addition, this research combines the K-mean clustering as a powerful tool for analyzing big data with TAOV that is a novel MCDM technique. Employing this integration would guarantee the quality of the segmentation. Likewise, a cluster analysis is performed jointly with the EDAS technique to optimize the size of the clusters. Finally, the CRM models and their combination with CLV are considered to focus more on the qualitative aspects of the analysis. The novel model combines these two tools with outputs derived from the Apriori algorithm to formulate short-term and long-term banking strategies.

3. Methodology

In this study, the actual data of 20,000 real accounts from customers operating on their accounts were selected from among a million customers by the studied bank. It should be noted that each customer has one or more real accounts because the information is classified according to the national customer code. The transactions of these clients are analyzed monthly over the past 24 months. Also, banking and information technology experts' opinions are used in various steps of this research to access and analyze banking and business information. The following is an overview of the implementation of this article.

Step 1: To prepare data, duplicate accounts, as well as empty or noisy fields, are refined after logging into Excel software. In the end, 20,000 data are selected as clean data for the survey. Preprocessing operations include processes such as correction or deletion of inappropriate data, determination of permissible limits, and correction of unauthorized values, recalculation, and deletion of data with the highest Standard deviation (SD). Statistical methods are then used to normalize the data x by Eq. (32) (Mahdiraji *et al.*, 2019).

$$\frac{x - \min}{\max - \min} \tag{32}$$

Step 2: Following the data preparation, in this step, a novel model found on five criteria is developed by banking experts.

(1) *Account Type (A_T):* This criterion specifies the type of customer account. The highest value is assigned to a checking account with a value of 9, followed by a short-term account with a value of 5, and finally a long-term account with a value of 1. It should be noted that the valuation of these scores is carried out in a separate study by the bank experts under the supervision of the planning and marketing department.

Since this criterion is nominal and each customer may have between 1 and 3 accounts in the bank, to extract the weights of its performance in all accounts and to make a value distinction between the types of accounts, they will be weighted after a single weight is multiplied by the other criteria. Hence, an adjusted weighting model is created, which is similar to the weighted RFM (WRFM) model in the previous literature. The pre-values are computed by Eq. (33).

$$\text{Pre} - \text{values} = \frac{\sum \text{ sum weights of each individual}'\text{s accounts}}{\sum \text{ sum weights of all accounts}} \tag{33}$$

Pre-valued accounts are weighed and multiplied by the cumulative state of the other criteria by experts. These values are normalized by the standard method (max–min). The average deposits of Type I to III deposits (checking, short term, long term) are multiplied as a single weight by other criteria, and the A_T criterion plays the role of weight here.

- If one customer has only one account type, the other two accounts are considered zero.
- The value of each individual's accounts is calculated by Eq. (34). It is then normalized linearly to obtain A_T.

$$A_T = \sum \text{value of checking accounts} + \sum \text{value of short term accounts} + \sum \text{value of long term accoumts} \quad (34)$$

- According to the values accrued to each account, the maximum possible value per person is 15, that is, it has all three accounts active. The minimum possible value is at least 1 in the sense that it has only one long-term account.

(2) *Average money (M):* This criterion represents the daily average of the minimum remaining at the end of the day in the real customer's account. It should be noted that the monthly interval is used to calculate the average. M is measured by Eq. (35).

$$M = \frac{\sum \text{Min of Remainings per day during a month}}{\text{number of month}'\text{s day}} \quad (35)$$

(3) *Average transaction frequency (F):* This criterion demonstrates the average number of customer transactions per month. F is obtained by Eq. (36).

$$F = \frac{\sum \text{number of transactions per day during a month}}{\text{number of month}'\text{s day}} \quad (36)$$

(4) *Average daily debt turnover (D_C):* This criterion speaks for the monthly average of money transferred from a customer's account to other banks. D_C is calculated by Eq. (37).

$$D_C = \frac{\sum \text{money transferred to other banks per day during a month}}{\text{number of month}'\text{s day}} \quad (37)$$

(5) *Average daily cash turnover (C_C):* This criterion introduces the monthly average of money transferred to customer accounts from other banks. C_C is attained by Eq. (38).

$$C_C = \frac{\sum \text{money transferred from other banks per day during a month}}{\text{number of month}'\text{s day}} \quad (38)$$

- The start and the end of the month are following the Iranian calendar. The year starts from the 22nd of May 2017 and ends on the 22nd of May 2019.
- All criteria are positive from the bank's point of view, which means that the larger the finance, the better the financial behavior of the customer. The only exception is C_C, which has a negative effect. The reason is that the customer has transferred the money to other banks. This is a detrimental performance from the bank's viewpoint.
- The average mean of the M, F, D_C, C_C criteria calculated for the last 24 months (calculated as the cumulative sum of each criterion over the number of months that is 24 months).

- *Step 3: K*-means clustering is performed for the value of *K* from 2 to 10. Cluster quality evaluation indices (SIH, DB and CH) are computed for each value of *K*. Then, the EDAS ranking method is implemented to select the optimal *K*, which is the optimal number of clusters.
- *Step 4:* In this step, the weights of the criteria are extracted by the BWM before ranking the clusters. The banking industry experts have developed a questionnaire to implement the BWM and the TAOV method. In this research ten experts participated in a panel including three from private banks, three from public banks, two academicians familiar with MCDM methods and clustering approaches, and two experts from the Central Bank of Iran. For the BWM , two questionnaires were used to compare the best criterion (most important criteria) with others and one questionnaire to compare other criteria with the worst criterion (least important criteria). These two questionnaires were completed during a one-hour face-to-face session after the briefing of the method that was presented by the research team. After the weight of the criterion is extracted, the cluster matrix is formed, and they are ranked using the TAOV method. For the TAOV method, the third questionnaire was used to complete the decision matrix and evaluate each customer cluster by each criterion. Afterward, the hidden rules of the clusters are excerpted by applying the associative rules and the Apriori algorithm. Finally, in the last step, strategy development is employed to improve CRM based on the concepts of customer lifetime value.

4. Analysis and results

In the first step to prepare the data, the data were obtained based on criteria in collaboration with banking experts. In this step, the information of 1,100,000 customers is provided for the research. Most of the provided data lacked criterion values and were unusable. As a consequence, simple sampling is based on a simple 5% value of (n/N), where (n) signifies the number of samples selected, and (N) is the total population number. Here, the information from 20,000 customers is gathered among all data for the data mining process. After selecting the data, in the next stage, the data is cleaned and merged. In this study, for the deletion of data, if the data distribution is between 3 and 5 times the SD in both positive and negative directions, they are identified as outliers. Moreover, if the data distribution is not normal, it is used to identify boxed graphs, with data ranging from 25 to 75% of the data being selected as the desired data and other values replaced with minimum and maximum points to obtain reliable results after clustering. Later, a model similar to the WRFM model is employed to integrate them. Since customer information is found on national code, and each customer may have multiple accounts, to facilitate the process of clustering with experts, a separate survey in the bank's marketing unit evaluates all accounts by weighing them as a single weight. Finally, the data are normalized by Eq. (32).

In the second step, the AFMDC [Account Type (A); Average transaction frequency (F); Average money (M); Average daily debt turnover (D); Average daily cash turnover (C)] model is formed, and in the third step, the optimal number of clusters is determined. For this purpose, three indices of the internal of the clusters' performance, including SIH, DC and CH, are considered as the basis of the work. *K*-means clustering is performed, and values 2 to 10 are calculated as the number of clusters. Then the internal quality indices are measured by Eqs. (2) to (12). A decision-making matrix is formed, which is eliminated in Table 3. In the following decision matrix, the EDAS technique is employed to find the optimal value of K. To apply this technique, the average distance is obtained by Eq. (13), and PDA and NDA are calculated by Eqs. (14) and (15) which are demonstrated in Table 3.

Therefore, the computation is performed by Eqs. (17) to (26), and the result is illustrated in Table 4. The results determined the optimal value of 6 for *K*. Thus, six clusters are selected for classifying the customers.

In step 4, K-mean clustering is accomplished for $k = 6$ and gathered data of 20,000 customers. The frequency of each cluster is displayed in Table 5. As can be seen from Table 5, Cluster 6 has the highest number of observations, accounting for 33.65% of the items. Cluster 4 has the lowest number of observations. The comparative results have been depicted in the pie chart of Figure 6. The other highlighted information which is attained by K-means is the clusters' centroids. This information is manifested in Table 5.

In consonance with Table 5,

(1) Concerning M, the centers of the two clusters 2 and 3 are very similar, and the most significant distinction is observed between the two clusters 2 and 6 for this criterion. Consequently, if clusters are analyzed by M, probably, similarities between the two

Table 3. The optimal number of clusters

	Decision matrix			PDA matrix			NDA matrix		
Number of clusters	SIH	CH	DB	SIH	CH	DB	SIH	CH	DB
2	0.624	17152.663	0.795	0	0	0	0.040	0.201	0.034
3	0.589	13644.855	0.720	0	0	0.063	0.093	0.364	0
4	0.652	21012.742	0.7474	0.002	0	0.028	0	0.021	0
5	0.661	24822.673	0.713	0.016	0.155	0.072	0	0	0
6	0.691	25261.984	0.711	0.062	0.176	0.074	0	0	0
7	0.700	24401.597	0.767	0.076	0.136	0.0021	0	0	0
8	0.666	24408.570	0.789	0.024	0.136	0	0	0	0.026
9	0.661	24659.096	0.828	0.016	0.148	0	0	0	0.076
10	0.608	17924.661	0.848	0	0	0	0.064	0.165	0.102

Table 4. The optimal value of K by EDAS

Number of clusters	SP	NSP	SN	NSN	AS
2	0	0	0.091	0.398	0.199
3	0.021	0.201	0.152	0	0.100
4	0.010	0.098	0.007	0.952	0.525
5	0.0814	0.779	0	1	0.889
6	0.104	1	0	1	1
7	0.071	0.685	0	1	0.842
8	0.0537	0.514	0.008	0.941	0.7280
9	0.054	0.525	0.025	0.832	0.679
10	0	0	0.110	0.274	0.137

Note(s): SP, Sum of Positive Distances; NSP, Normalized Sum of Positive; SN, Sum of Negative Distances; NSN, Normalized Sum of Negative; AS, Attractiveness Score

Table 5. Clusters' size and centroids

	Size			Centroids			
Cluster number	Item frequencies	Item relative frequencies percentage	Rank of size	M	F	D_C	C_C
1	3,562	17.81	3	0.695	0.901	0.488	0.516
2	4,269	21.34	2	0.933	0.951	0.038	0.962
3	1,922	9.61	5	0.938	0.448	0.091	0.916
4	1,341	6.70	6	0.372	0.730	0.105	0.882
5	2,176	10.88	4	0.677	0.414	0.587	0.404
6	6,730	33.65	1	0.027	0.018	0.977	0.019

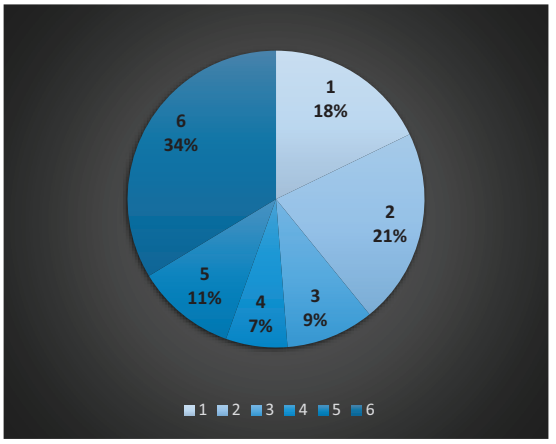


Figure 6.
Pie chart of clusters' size

clusters 2 and 3 and significant differences between the two clusters 2 and 6 will be found, which is further confirmed.

(2) Regarding F, centers of two clusters 1 and 2 have the most similarity, and centers of two clusters 2 and 6 also have the most difference.

(3) Regarding C_C, the centers of the two clusters 2 and 3 have the most similarities, and the centers of the two clusters 2 and 6 have the most differences.

(4) Taking into account D_C, the centers of the two clusters 1 and 5 have the most similarities, and the centers of the two clusters 3 and 6 have the most differences. Figure 7 depicts the demographic information of clusters.

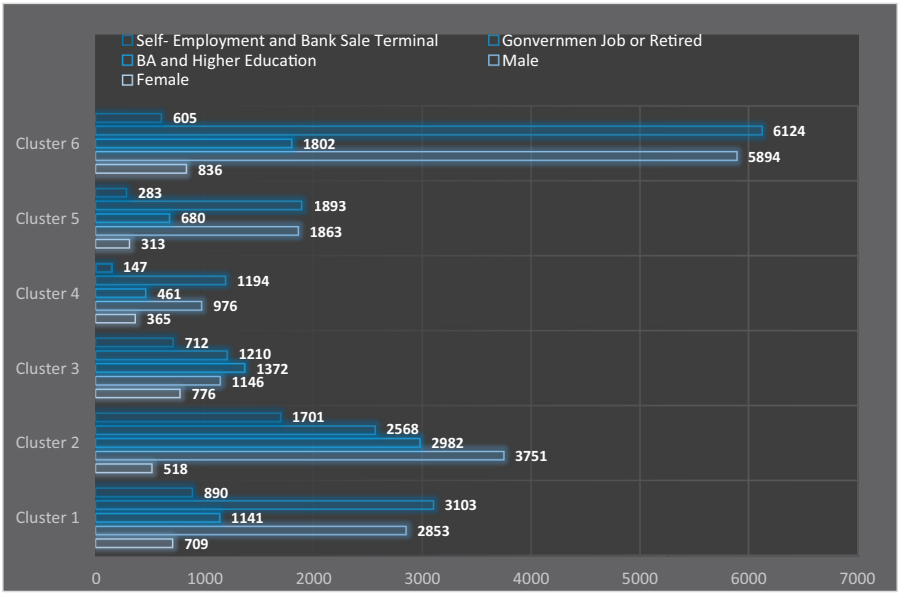


Figure 7.
Demographic information for clusters

In the fifth step, after forming the clusters, the weights of the criteria are extracted by the BWM. The average monetary (AVG-*M*) is determined as the best criterion for evaluating a customer, according to banking experts. This is because having more money in an account makes the bank more profitable by monetization. On the other hand, the SD of the daily creditor turnover (STDEV-C_c) is the worst criterion from the banking experts' point of view. Other criteria contain average daily creditor turnover (AVG-C_c), average daily frequency (AVG-*F*) and SD of monetary (STDEV-*M*). Next, the best and worst criterion are compared with other criteria and assigned numbers between 1 and 9. The number 1 represents the same value of the criteria, and 9 represents the highest priority. The results are demonstrated in Table 6. To obtain the weights by the BWM, the model of Eq. (13) is constructed by the data of Table 6 and solved. Table 6 shows the weight of each criterion.

After extracting the weight of the criteria in step 5, the clusters are ranked by the TAOV method. In Phase I, the decision matrix is formed, which is illustrated in Table 7.

The decision matrix is normalized by Eq. (24) for beneficial criteria and Eq. (25) for cost criteria. Thence it is weighted by applying Eq. (26). Table 8 demonstrates the weighted normalized matrix. Accordingly, the equivalent matrix is built as described by Eqs. (27) to (29) and is shown in Table 8. The value of total attractiveness (TA) and normalized total attractiveness (NTA) is computed by Eqs. (30) and (31). The result of prioritization by TAOV is depicted in Table 8.

5. Practical implications

After ranking the clusters, rules are extracted, applying the Apriori algorithm. In this research, the minimum support is determined by 2%, and the minimum confidence is 70%. The lift criterion, which is employed to evaluate associative rules, is obtained by dividing the degree of confidence by the support. Any higher value than 1 indicates the attractiveness of the rule. The clusters' rules are presented in Table 9.

Cluster 1 rules: The most remarkable correlation found in this cluster is the significant relationship between the two indices of daily creditor turnover (C_C) and daily debtor turnover (D_C). The results of these rules are presented in Table 9a.

Table 6. BWM paired comparisons and weights

Criteria	Comparison		Criteria weights
	Best criterion	Worst criterion	
AVG *M*	1	9	0.45
AVG C_C	3	6	0.29
AVG *F*	5	5	0.14
ST DEV *M*	7	4	0.08
ST DEV C_C	9	1	0.04

Table 7. TAOV decision matrix

Cluster	AVG *M*	AVG C_c	AVG *F*	ST DEV *M*	ST DEV C_c
1	0.695	0.516	0.901	0.205	0.120
2	0.933	0.962	0.951	0.113	0.079
3	0.938	0.916	0.448	0.119	0.125
4	0.372	0.882	0.730	0.168	0.135
5	0.677	0.404	0.414	0.251	0.150
6	0.027	0.019	0.018	0.078	0.065
Criteria type	*B*	*B*	*B*	*C*	*C*

Note(s): *B*: benefit criteria, *C*: cost criteria

Cluster	Weighted normalized matrix						Equivalent matrix							
	AVG M	AVG C_c	AVG F	ST DEV M	ST DEV C_c	AVG M	AVG C_c	AVG F	ST DEV M	ST DEV C_c	TA	NTA	Rank
1	0.331	0.154	0.132	0.031	0.022	0.455	0.258	−0.066	0.068	0.002	0.9532	0.175	3
2	0.444	0.288	0.139	0.057	0.033	0.630	0.383	−0.119	0.0600	−0.001	1.333	0.245	1
3	0.447	0.274	0.065	0.055	0.021	0.572	0.346	−0.156	0.0603	−0.003	1.277	0.235	2
4	0.177	0.264	0.107	0.038	0.019	0.394	0.252	−0.043	−0.026	−0.003	0.803	0.148	5
5	0.322	0.121	0.060	0.026	0.017	0.372	0.210	−0.093	0.073	0.000	0.851	0.156	4
6	0.013	0.005	0.002	0.083	0.040	−0.080	0.079	−0.009	0.0008	0.003	0.206	0.038	6

Table 8.
TA and NA calculation

9a. Cluster 1 rules				
Lift	Confidence	Support	Antecedent	Consequent
41	82%	2%	D_c	C_c

9b. Cluster 2 rules				
Lift	Confidence	Support	Antecedent	Consequent
47.5	95%	2%	D_c	C_c
38	76%	2%	C_c	M
36.5	73%	2%	D_c and C_c	F

9c. Cluster 3 rules				
Lift	Confidence	Support	Antecedent	Consequent
45.5	89%	2%	D_C	C_C
39	78%	2%	C_C	M

9d. Cluster 4 rules				
Lift	Confidence	Support	Antecedent	Consequent
48	96%	2%	C_C	D_C
38	76%	2%	F	D_C
37.5	75%	2%	D_C	M
35.5	71%	2%	F	M

9e. Cluster 5 rules				
Lift	Confidence	Support	Antecedent	Consequent
37	74%	2%	F	D_C
36.5	73%	2%	C_C	D_C
35	70%	2%	F	M

9f. Cluster 6 rules				
Lift	Confidence	Support	Antecedent	Consequent
45.5	99%	2%	C_c	D_C
48	96%	2%	M	F
45	90%	2%	M	D_C

Table 9.
Clusters' rules

The level of education in this cluster is relatively similar to that of clusters 4, 5 and 6 and is not high. In this cluster, the daily creditor turnover (C_C) is the result of the daily debtor turnover (D_C). The interpretation of this pattern may be related to the fact that whenever money is withdrawn from one of the accounts in this cluster, it is subsequently deposited into this account or possibly into other accounts in the bank under study. Moreover, 80% of the people in this cluster are men. One of the reasons that the number of men in this bank is higher than other banks is that this bank acts as the operating bank to pay the salaries of a large government agency, and often employees and retirees are men. Furthermore, 24% of the people in this cluster are self-employed, meaning that about one-fourth of the people in the bank are outside of those with whom the bank has a paying relationship. One of the plans the bank has offered to its customers in marketing is to purchase a commodity loan plan, whereby business owners and sellers sell their goods through bank credit, and the bank receives some of the sellers' profits and loan repayments. The need for such a mechanism is to

open an account by sellers and customers so that business owners and sellers can fall into this cluster of 24% of the population.

Cluster 2 rules: The most highlighted correlation discovered in this cluster, which is also the most important cluster, is the relationship between the two indices of the daily debtor (D_C) and creditor turnover (C_C). Moreover, a significant relationship between the two indices of average monetary (M) and daily creditor turnover (C_C), the average frequency of transactions (F) with daily debtor turnover (D_C) and daily creditor turnover (Cc) is found. The results of these rules are given in Table 9b.

Approximately 70% of people in this cluster have a high level of education. Most of them are senior or senior bank employees with relatively higher salaries of the organization that the bank is responsible for paying. About 40% of the people in this cluster are self-employed, other than employees. These people are generally sellers who have used a point-of-sale terminal (POS) machine from the bank to borrow money and provide this money to customers. In this cluster, the daily creditor turnover (C_C) is the result of the daily debtor turnover (D_C). The explanation of this set implies that whenever money is withdrawn from one of the accounts in the cluster, the money is subsequently deposited into this account or possibly into other accounts. For instance, whenever money is withdrawn from one of the sellers' accounts, a significant amount is deposited into their accounts because it is a bank loan repayment that is deposited into one of the sellers' accounts as a deduction. Here, the sellers have two types of accounts that are nondeductible until the confirmation of the purchase of the goods by bank and the short-term account, which is available after 72 h. The average amount of money (M) is the result of the daily turnover items (C_C), which means that when the money is deposited into the cluster (from other banks), their average money has increased. The average frequency of transactions (F) is also related to the average daily debit (D_C) turnover, proposing that numerous transactions cause the money to be withdrawn from their accounts and transferred to other banks. As with other clusters, males are the predominant population, and most of the staff in the organization are males, according to a report previously obtained from the bank. Besides, most vendors and business owners in the cluster are male.

Cluster 3 rules: In the third cluster, the relationship between the monetary (M) and the daily creditor turnover (C_C) as well as the relationship between the daily creditor turnover (C_C) and the daily debtor turnover (D_C) has been treasured. The results of these rules are proposed in Table 9c.

This cluster, identified as the second top cluster, has the same characteristics as the second cluster. Thus, the daily creditor turnover (C_C) is the result of the daily debtor turnover (D_C). Those in this cluster are also well-educated, with about 71% having a bachelor's degree or higher. Men are the dominant population in this cluster. About 37% of the people in the cluster are self-employed, reflecting the presence of a significant population of sellers and contractors in the bank and purchase lending. In this cluster, whenever the account holder transfers money from other banks to the bank under study [daily creditor turnover (C_C)], the average amount of money (M) is increased. Hence, the money that remains in the person's account and in other accounts or banks is no longer transferred, and this factor is one of the important factors that has increased the value of this cluster compared to other clusters. This is due to this issue that the monetary value has a higher weight than other factors and indicators.

Cluster 4 rules: In this cluster, there are frequent patterns between the daily creditor turnover (C_C) and the daily debit turnover (D_C), and so on between the average transaction frequency (F) and the daily debtor turnover (D_C). Moreover, a pattern is found between the daily debit turnover (D_C) and the average monetary value (M) and again the same factor with the average frequency of transactions (F). The results of these rules are presented in Table 9d.

In this cluster, the daily debtor turnover (D_C) is the result of the daily creditor turnover (C_C) items, meaning that whenever the money is deposited into other people's accounts or other accounts, the money is transferred out of the person's account and transferred to other banks. The dominant population is men. The level of education is not relatively high and this is probably due to the relatively lower remuneration of those present in the cluster compared to the other people mentioned in the contract with the bank. Daily debtor turnover (D_C) is the result of the average frequency of transactions (F). This signifies that the money has been withdrawn from the customer's account and transferred to the outside bank, which is a negative factor for customer ratings. The average amount of money (M) is the result of the daily creditor turnover (C_C). This represents that when the money is deposited into the accounts of the people in this cluster, their average money (M) is significant. Besides, the average frequency of transactions (F) in this cluster resulted in a significant monetary average (M).

Cluster 5 rules: In this cluster, there is a recurring pattern between the daily debtor turnover (D_C) and the frequency of transactions (F) as well as the daily debtor turnover (D_C) and the daily creditor turnover (C_C), and the average monetary value (M) and the frequency of transactions (F). The results of these rules are given in Table 9e.

In this cluster, 31% of people have a bachelor's degree or higher. The majority of the population is still male. The daily debtor turnover (D_C) is the result of the daily creditor turnover (C_C), which means that whenever money is deposited into these accounts, it is seen that it has been transferred to other accounts or banks. The daily debtor turnover (D_C) is associated with the average frequency of transactions (F), meaning that whenever the transactions are significant and volatile, the money is transferred to other accounts or banks from those in this cluster. Furthermore, the average monetary value (M) is shown by the mean frequency of transactions (F). Remark that 13% of the people in this cluster are self-employed business owners, which eliminates that there are fewer people involved in the bank's marketing campaigns and fewer people outside the bank.

Cluster 6 rules: Cluster 6, which has the worst performance in terms of financial behavior, has recurring patterns between the average monetary value (M) and the daily debtor turnover (D_C), the transaction frequency (F) and the average monetary value (M), the daily debtor turnover (D_C) and daily creditor turnover (C_C). The results of these rules are given in Table 9f.

This cluster, which has the lowest rank among other clusters, has the highest number of customers, with 27% having a bachelor's or higher level of education, indicating that the people contracted with the bank in this cluster are in the bottom ranks of organizations. The dominant population is men. Here, the daily debtor turnover (D_C) is the result of the daily creditor turnover (C_C), which conveys that whenever money is deposited into these accounts, it is transferred to other accounts or banks. Note that 9% of the people are self-employed business owners, which is still the least involved in the bank's marketing campaigns, and fewer people outside the bank are attracted to the bank's marketing plans. The daily debtor turnover (D_C) is the result of the average amount of money (M), and whenever a person is making a change in their average amount of money (M), he is withdrawing money from his account to other accounts or banks. In this cluster, the average frequency of transactions (F) is the result of the average value of money (M), which is probably the worst performing cluster in terms of performance. Thus, when the average value of money (M) is decreasing, one is performing frequent transactions to transfer money to other accounts or banks.

In the final step, the strategy for interacting with clusters is formulated with the inspiration of the CLV. Customer value refers to the potential interaction of customers with the industry over specific periods. Once the industry understands customer value and realizes that customer value can deliver customized service to different customers, then CRM is achieved effectively. There are generally four steps in the customer life cycle as follows:

(1) *Potential customers:* People who are not yet customers but are targeted in the market.

(2) *Reacting customer:* potential customers who are interested in and respond to a product or service.

(3) *Active customers:* people who currently use a product or service from the organization.

(4) *Former customers:* such people are not good customers because they have not been targeted for a long time and have moved their purchases to compete with products.

Table 10 provides some suggested strategies for interacting with customers based on the characteristics of the individuals presented in each cluster.

Cluster	Rank	Customer type	Characteristics	Strategies
1	3	High-interacting actual customers	(1) People with moderate education (2) Appropriate creditor turnover (3) Business interactions with the bank	(1) 24/7 customer service (2) Integrated organization of customer accounts (3) Marketing for business people through advertising, brochures, and informative and persuasive ads (4) Offering new banking services at a reasonable fee
2	1	Extremely loyal customers	(1) Educated people (2) High creditor turnover (3) Moderate monetary value (4) More business interactions with other banks than other clusters	(1) Digital and self-service branches (2) 24/7 customer service (3) Developing electronic wallet services (using mobile as a bank card) (4) Development of communication services and interactions with foreign banks (5) Providing attractive services such as insurance, guarantees and lending facilities to loyal customers (6) Case rewards (7) Delegating branches to handle such customers (8) Marketing for business people through advertisements, brochures and informative reminders (9) Offering new banking services at a reasonable fee (10) Increasing the return on equity (11) Providing financial reports (12) Launching an online inquiry on the Central Bank Portal of the bank network for customers to access easily and rapidly the status of return checks and customer facilities (13) Organizing festival of sales terminals (14) Organizing e-service festivals

(*continued*)

Table 10.
Suggested strategies for clusters

Cluster	Rank	Customer type	Characteristics	Strategies
3	2	Loyal customers	(1) Educated people (2) High creditor turnover (3) More business interactions with other banks than other clusters	(1) Creating special payment facility conditions (2) Case rewards (3) 24/7 customer service (4) Developing electronic wallet services (using mobile as a bank card) (5) Development of communication services and interactions with foreign banks (6) Providing attractive services such as insurance, guarantees and lending facilities to loyal customers (7) Offering new banking services at a reasonable fee (8) Marketing for business people through reminders, brochures and advertisements
4	5	Low-interacting potential customers	(1) People with moderate education (2) High debtor turnover	(1) Sending bank brochures and services (2) Announcing and holding celebrations, lotteries and advertisements
5	4	Moderate-interacting potential customers	(1) People with moderate education (2) High debtor turnover (3) Low monetary value	(1) Sending bank brochures, services, encouraging and informative ads (2) Holding ceremonies at festivals, sweepstakes, meetings and exhibitions
6	6	Missing customers	(1) People with moderate education (2) Low monetary average (3) High debtor turnover (4) Quite low business interactions	(1) Proposing encouraging and informative television advertising (2) Accelerating the provision of short-term incentive services to attract customers and hold rallies and raffles (3) Providing personal information via mobile or other means of communication such as post (4) Offering attractive profits by introducing bank investment funds in stock exchanges and businesses (5) Facilitating communication channel processes such as mobile banking and internet banking

Table 10.

Some important guidelines for CRM and CLV enhancement are as follows:

(1) making the most of employees' ability to generate growth and a sense of belonging to the organization to form compassionate interaction with customers;

(2) concentrating on developing employees' abilities to improve their performance in banking processes to expedite client-side work:

(3) holding communication skills training courses for the staff;

(4) creating relationships between managers and employees to build a constructive relationship and create a positive spirit in employees and enhancing their productivity and attracting customers;

(5) paying more attention to the physical and hygienic environment and customer requirements in physical communication channels;

(6) creating an autonomous relationship for customers to create a sense of belonging to the organization;

(7) congratulating customers through brochures and other low-cost communication channels such as mobile banking and e-mail;

(8) promoting the level of vitality of the staff by formulating the organized plans;

(9) designing a performance appraisal system with a staff performance management approach;

(10) establishing a merit-based appointment system.

6. Conclusion

In this study, a multi-attribute data mining model was used to segment the customers and group them into six clusters by analyzing 20,000 customer records in the financial services industry. Clusters 2, 3, 1, 5, 4 and 6 have the highest to lowest values, respectively. The Apriori algorithm was employed next to extract frequent patterns of customer financial behavior. Demographic characteristics and financial transactions of customers were among the factors analyzed in this study. As a result, the following six customer types of highly loyal, loyal, high-interacting, high-interacting, low-interacting and missing customers were identified. Appropriate strategies for interacting with each customer type were proposed based on the opinions of banking experts and the literature reviewed regarding customer management systems and their lifetime value.

Theoretically, although relevant articles in the same area have implemented clustering methods such as *K*-mean to categorize customers in different groups (e.g. Anitha and Patil, 2019; Basak *et al.*, 2019), integrating the categorized group of customers and then ranking them based on specific criteria and extracting rules and plans to deal with each cluster has not been previously investigated in an integrated fashion. Moreover, besides other applications of the BWM and the TAOV method in manufacturing, construction, etc. decision-making problems (e.g. Amoozad Mahdiraji *et al.*, 2018; Hajiagha *et al.*, 2018; Mahdiraji *et al.*, 2019), in this research, a new application of these MCDM methods in the banking industry and CLV analysis has been designed and scheduled in a novel approach. Practically, managers in financial organizations and especially in the banking industry can benefit from the results of this research and the integrated decision-making and data mining approach proposed in this study. Combining different analytical tools can benefit banks categorize their customers and, as a result, design suitable marketing strategies for each group. Accordingly, service-oriented organizations can allocate their budget, plans, time, etc. more optimally toward increasing the satisfaction of each cluster of customers based on their value and preferences.

There are some limitations regarding the BWM and the EDAS, and TAOV methods utilized in this research. First, the methods used in this research were all based on crisp numbers and the assumption of specific situations. In problems involving uncertainties, interval, grey, fuzzy, interval fuzzy, intuitionistic fuzzy or hesitant fuzzy numbers and methods could be used to deal with ambiguity and uncertainty in real-world problems. Considering the uncertain circumstances in today's market, adopting uncertain approaches in quantitative methods produces more reasonable outputs. From the methodological

perspective, other weighing methods, such as the stepwise weight assessment ratio analysis (SWARA) or the SECA method, could be used in conjunction with other alternative ranking methods such as the technique for order of preference by similarity to ideal solution (TOPSIS) or AHP to establish some benchmarks for data analysis and interpretation. Moreover, the authors implemented the *K*-means method for clustering the customers in this research; however, as indicated in Figure 4, other methods are also applicable to compare and benchmark the clustering results instead of SIH, CH and DB criteria for identifying the most appropriate number of clusters. Furthermore, the results of this research are based on limited data obtained from 20,000 customer records from the banking industry in an emerging economy. Researchers can focus on similar approaches by employing big data in future studies in other industries, sectors, regions or countries to generalize the results and present more inclusive implications.

References

Alizadeh Zoeram, A. and Karimi Mazidi, A.R. (2018), "New approach for customer clustering by integrating the LRFM model and fuzzy inference system", *Iranian Journal of Management Studies*, Vol. 11 No. 2, pp. 351-378.

Amoozad Mahdiraji, H., Arzaghi, S., Stauskis, G. and Zavadskas, E.K. (2018), "A hybrid fuzzy BWM-COPRAS method for analyzing key factors of sustainable architecture", *Sustainability*, Vol. 10 No. 5, pp. 16-26.

Anitha, P. and Patil, M.M. (2019), "RFM model for customer purchase behavior using K-means algorithm", *Journal of King Saud University – Computer and Information Sciences*, doi: 10.1016/j.jksuci.2019.12.011.

Aryuni, M. and Miranda, E., (2018), "Customer segmentation in XYZ bank using K-means and K-medoids clustering", *2018 International Conference on Information Management and Technology (ICIMTech)*, September, pp. 1-9.confproc

Ballestar, M.T., Grau-Carles, P. and Sainz, J. (2018), "Customer segmentation in E-commerce: applications to the cashback business model", *Journal of Business Research*, Vol. 88, pp. 407-414.

Basak, A., Sinha, A., Dey, R. and Shaw, K. (2019), "Prediction future disaster using convex hull K-mean, an approach", *IEMECON 2019 – 9th Annual Information Technology, Electromechanical Engineering, and Microelectronics Conference*, Institute of Electrical and Electronics Engineers, pp. 237-241.

Bhambri, V. (2011), "Application of data mining in banking sector", *International Journal of Computer Science and Technology*, Vol. 2 No. 2, pp. 199-202.

Birant, D. (2011), "Data mining using RFM analysis", in *Knowledge-Oriented Applications in Data Mining*.

De Caigny, A., Coussement, K. and De Bock, K.W. (2018), "A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees", *European Journal of Operational Research*, Vol. 269 No. 2, pp. 760-772.

Calinski, T. and Harabasz, J. (1974), "A dendrite method for cluster analysis", *Communications in Statistics – Theory and Methods*, Vol. 3 No. 1, pp. 1-27.

Casado, R. and Younas, M. (2015), "Emerging trends and technologies in big data processing", *Concurrency and Computation: Practice and Experience*, Vol. 27 No. 8, pp. 2078-2091.

Chen, C.P. and Zhang, C.Y. (2014), "Data-intensive applications, challenges, techniques and technologies: a survey on big data", *Information Sciences*, Vol. 275, pp. 314-347.

Cheng, C.H. and Chen, Y.S. (2009), "Classifying the segmentation of customer value via RFM model and RS theory", *Expert Systems with Applications*, Vol. 36 No. 3, pp. 4176-4184.

Cheung, Y.M. (2003), "K-means: a new generalized k-means clustering algorithm", *Pattern Recognition Letters*, Vol. 24 No. 15, pp. 2883-2893.

Chiang, W.Y. (2017), "Discovering customer value for marketing systems: an empirical case study", *International Journal of Production Research*, Vol. 55 No. 17, pp. 5157-5167.

Chiu, S. and Tavella, D. (2008), *Data Mining and Market Intelligence for Optimal Marketing Returns*, Routledge, London.

Çınar, K., Yetimoğlu, S. and Kaplan, U. (2020), *The Role of Market Segmentation and Target Marketing Strategies to Increase Occupancy Rates and Sales Opportunities of Hotel Enterprises*, Springer, Cham, pp. 521-528.

Dachyar, M., Esperanca, F.M. and Nurcahyo, R. (2019), "Loyalty improvement of Indonesian local brand fashion customer based on customer lifetime value (CLV) segmentation", *IOP Conference Series: Materials Science and Engineering*, Vol. 598, Institute of Physics Publishing.

Davidson, I. (2002), "Understanding K-means non-hierarchical clustering", *SUNY Albany Technical Report*, Vol. 2, pp. 2-14.

Davies, D.L. and Bouldin, D.W. (1979), "A cluster separation measure", *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-1*, No. 2, pp. 224-227.

De Marco, M., Fantozzi, P., Fornaro, C., Laura, L. and Miloso, A. (2021), "Cognitive analytics management of the customer lifetime value: an artificial neural network approach", *Journal of Enterprise Information Management*, Vol. 34 No. 2, pp. 679-696.

Dibb, S. (1998), "Market segmentation: strategies for success", *Marketing Intelligence and Planning*, Vol. 16 No. 7, pp. 394-406.

Dursun, A. and Caber, M. (2016), "Using data mining techniques for profiling profitable hotel customers: an application of RFM analysis", *Tourism Management Perspectives*, Vol. 18, pp. 153-160.

Ekbia, H., Mattioli, M., Kouper, I., Arave, G., Ghazinejad, A., Bowman, T., Suri, V.R., Tsou, A., Weingart, S. and Sugimoto, C.R. (2015), "Big data, bigger dilemmas: a critical review", *Journal of the Association for Information Science and Technology*, Vol. 66 No. 8, pp. 1523-1545.

Estrella-Ramón, A., Sánchez-Pérez, M., Swinnen, G. and VanHoof, K. (2017), "A model to improve management of banking customers", *Industrial Management and Data Systems*, Vol. 117 No. 2, pp. 250-266.

Fader, P.S., Hardie, B.G.S. and Lee, K.L., (2005), "RFM and CLV: using ISO-value curves for customer base analysis", *Journal of Marketing Research*, Vol. 42, No. 4, pp. 415-430.

Farajian, M.A. and Mohammadi, S. (2010), "Mining the banking customer behavior using clustering and association rules methods", *International Journal of Industrial Engineering and Production Research*, Vol. 21, No. 4, pp. 239-245.

Ghorabaee, M.K., Amiri, M., Zavadskas, E.K. and Antuchevičienė, J. (2017a), "Assessment of third-party logistics providers using a CRITIC–WASPAS approach with interval type-2 fuzzy sets", *Transport*, Vol. 32 No. 1, pp. 66-78.

Ghorabaee, M.K., Amiri, M., Zavadskas, E.K., Hooshmand, R. and Antuchevičienė, J. (2017b), "Fuzzy extension of the CODAS method for multi-criteria market segment evaluation", *Journal of Business Economics and Management*, Vol. 18 No. 1, pp. 1-19.

Gireesha, O., Somu, N., Raman, M.G., Reddy, M.S., Kirthivasan, K. and Sriram, V.S. (2020), "WNN-EDAS: a wavelet neural network based multi-criteria decision-making approach for cloud service selection", in *Computational Intelligence in Pattern Recognition*, Springer, Singapore, pp. 853-865.

Gobble, M.A.M. (2013), "Big data: the next big thing in innovation", *Research-Technology Management*, Vol. 56 No. 1, pp. 64-67.

Gupta, K., Goyal, N. and Khatter, H. (2019), "Optimal reduction of noise in image processing using collaborative inpainting filtering with pillar K-mean clustering", *The Imaging Science Journal*, Vol. 67 No. 2, pp. 100-114.

Hajiagha, S.H.R., Mahdiraji, H.A. and Hashemi, S.S. (2018), "Total area based on orthogonal vectors (TAOV) as a novel method of multi-criteria decision aid", *Technological and Economic Development of Economy*, Vol. 24 No. 4, pp. 1679-1694.

He, Y., Lei, F., Wei, G., Wang, R., Wu, J. and Wei, C. (2019), "EDAS method for multiple attribute group decision making with probabilistic uncertain linguistic information and its application to green supplier selection", *International Journal of Computational Intelligence Systems*, Vol. 12 No. 2, pp. 1361-1370.

Hilbert, M. and López, P. (2011), "The world's technological capacity to store, communicate, and compute information", *Science*, Vol. 332 No. 6025, pp. 60-65.

Hu, X., Shi, Z., Yang, Y. and Chen, L. (2020), "Classification method of internet catering customer based on improved RFM model and cluster analysis", *2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*, April, IEEE, pp. 28-31.

Jacobs, A. (2009), "The pathologies of big data", *Communications of the ACM*, Vol. 52 No. 8, pp. 36-44.

Jagani, K., Oza, F.V. and Chauhan, H. (2020), "Customer segmentation and factors affecting willingness to order private label brands: an E-grocery shopper's perspective", in *Improving Marketing Strategies for Private Label Products*, IGI Global, pp. 227-253.

Jain, A.K. (2010), "Data clustering: 50 years beyond K-means", *Pattern Recognition Letters*, Vol. 31 No. 8, pp. 651-666.

Jain, A.K., Murty, M.N. and Flynn, P.J. (1999), "Data clustering: a review", *ACM Computing Surveys*, Vol. 31, pp. 264-323.

Khajvand, M., Zolfaghar, K., Ashoori, S. and Alizadeh, S. (2011), "Estimating customer lifetime value based on RFM analysis of customer purchase behavior: case study", *Procedia Computer Science*, Vol. 3, pp. 57-63.

Khan, S.M., Kharade, R.S., Lavange, V.S. and Pohare, D.B. (2019), "Detecting brain tumor using K-mean clustering and morphological operations", *IRJET*, Vol. 6, pp. 870-874.

Kimiagari, S., Keivanpour, S. and Haverila, M. (2021), "Developing a high-performance clustering framework for global market segmentation and strategic profiling", *Journal of Strategic Marketing*, Vol. 29 No. 2, pp. 93-116.

Kraska, T. (2013), "Finding the needle in the big data systems haystack", *IEEE internet Computing*, Vol. 17 No. 1, pp. 84-86.

Kuo, R.J., Amornnikun, P. and Nguyen, T.P.Q. (2020), "Metaheuristic-based possibilistic multivariate fuzzy weighted c-means algorithms for market segmentation", *Applied Soft Computing*, Vol. 96, 106639.

Leverin, A. and Liljander, V. (2006), "Does relationship marketing improve customer relationship satisfaction and loyalty?", *International Journal of Bank Marketing*, Vol. 24 No. 4, pp. 232-251.

Li, H., Yang, X., Xia, Y., Zheng, L., Yang, G. and Lv, P. (2018), "K-LRFMD: method of customer value segmentation in shared transportation filed based on improved K-means algorithm", *Journal of Physics: Conference Series*, IOP Publishing, Vol. 1060 No. 1, 012012.

Lumsden, S.A., Beldona, S. and Morrison, A.M. (2008), "Customer value in an all-inclusive travel vacation club: an application of the RFM framework", *Journal of Hospitality and Leisure Marketing*, Vol. 16 No. 3, pp. 270-285.

MacQueen, J., (1967), "Some methods for classification and analysis of multivariate observations", *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, June, Vol. 1, No. 14, pp. 281-297.confproc

Maghsoodi, A.I., Rasoulipanah, H., López, L.M., Liao, H. and Zavadskas, E.K. (2020), "Integrating interval-valued multi-granular 2-tuple linguistic BWM-CODAS approach with target-based attributes: site selection for a construction project", *Computers and Industrial Engineering*, Vol. 139, 106147.

Mahdiraji, H.A., Kazimieras Zavadskas, E., Kazeminia, A. and Abbasi Kamardi, A. (2019), "Marketing strategies evaluation based on big data analysis: a CLUSTERING-MCDM approach", *Economic research-Ekonomska istraživanja*, Vol. 32 No. 1, pp. 2882-2892.

Maji, G., Dutta, L. and Sen, S. (2019), "Targeted marketing and market share analysis on POS payment data using DW and OLAP", in *Emerging Technologies in Data Mining and Information Security*, Springer, Singapore, pp. 189-199.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and Hung Byers, A. (2011), *Big Data: The Next Frontier for Innovation, Competition, and Productivity*, McKinsey Global Institute, Chicago.

Miller, H.G. and Mork, P. (2013), "From data to decisions: a value chain for big data", *It Professional*, Vol. 15 No. 1, pp. 57-59.

Motlagh, O., Berry, A. and O'Neil, L. (2019), "Clustering of residential electricity customers using load time series", *Applied Energy*, Vol. 237, pp. 11-24.

Munusamy, S. and Murugesan, P. (2020), "Modified dynamic fuzzy c-means clustering algorithm–application in dynamic customer segmentation", *Applied Intelligence*, Vol. 50, pp. 1922-1942, doi: 10.1007/s10489-019-01626-x.

Newell, F. (1997), *The New Rules of Marketing: How to Use One-To-One Relationship Marketing to Be the Leader in Your Industry*, Irwin Professional Publishing, Oklahoma.

Newton, H. (2004), *Newton's Telecom Dictionary*, CMP Books, Kansas, p. 617.

Öner, S.C. and Öztayşi, B. (2017), "An interval valued hesitant fuzzy clustering approach for location clustering and customer segmentation", in *Advances in Fuzzy Logic and Technology 2017*, Springer, Cham, pp. 56-70.

Öztayşi, B., Sezgin, S. and Özok, A.F. (2011), "A measurement tool for customer relationship management processes", *Industrial Management and Data Systems*, Vol. 111 No. 6, pp. 943-960.

Parikh, Y. and Abdelfattah, E. (2020), "Clustering algorithms and RFM analysis performed on retail transactions", *2020 11th IEEE Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON)*, October, IEEE, pp. 506-511.

Peker, S., Kocyigit, A. and Eren, P.E. (2017), "LRFMP model for customer segmentation in the grocery retail industry: a case study", *Marketing Intelligence and Planning*, Vol. 35 No. 4, pp. 544-559.

Phan, T.C., Rieger, M.O. and Wang, M. (2019), "Segmentation of financial clients by attitudes and behavior", *International Journal of Bank Marketing*, Vol. 37 No. 1, pp. 44-68.

Qadadeh, W. and Abdallah, S. (2018), "Customers segmentation in the insurance company (TIC) dataset", *Procedia Computer Science*, Vol. 144, pp. 277-290.

Rahmadianti, R., Dhini, A. and Laoh, E. (2020), "Estimating customer lifetime value using LRFM model in pharmaceutical and medical device distribution company", *2020 International Conference on ICT for Smart Society (ICISS)*, November, IEEE, pp. 1-5.

Reichstein, T. and Salter, A. (2006), "Investigating the sources of process innovation among UK manufacturing firms", *Industrial and Corporate Change*, Vol. 15 No. 4, pp. 653-682.

Reutterer, T., Platzer, M. and Schröder, N. (2020), "Leveraging purchase regularity for predicting customer behavior the easy way", *International Journal of Research in Marketing*, Vol. 38 No. 1, pp. 194-215.

Rezaei, J. (2015), "Best-worst multi-criteria decision-making method", *Omega*, Vol. 53, pp. 49-57.

Rousseeuw, P.J. (1987), "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis", *Journal of Computational and Applied Mathematics*, Vol. 20, pp. 53-65.

Saeedi, M. and Albadvi, A. (2018), "A new 3D value model for customer segmentation: complex network approach", in *Applications of Data Management and Analysis*, Springer, Cham, pp. 129-145.

Safari, F., Safari, N. and Montazer, G.A. (2016), "Customer lifetime value determination based on RFM model", *Marketing Intelligence and Planning*, Vol. 34 No. 4, pp. 446-461.

Sarstedt, M. and Mooi, E. (2014), *A Concise Guide to Market Research. The Process, Data, and Methods Using IBM SPSS Statistics*, Springer, Berlin.

Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O.P., Tiwari, A., Er, M.J., Ding, W. and Lin, C.T. (2017), "A review of clustering techniques and developments", *Neurocomputing*, Vol. 267, pp. 664-681.

Sivasankar, E. and Vijaya, J. (2017), "Customer segmentation by various clustering approaches and building an effective hybrid learning system on churn prediction dataset", in *Advances in Intelligent Systems and Computing*, Springer Verlag, Vol. 556, pp. 181-91.

Smith, W.R. (1956), "Product differentiation and market segmentation as alternative marketing strategies", *Journal of Marketing*, Vol. 21 No. 1, pp. 3-8.

Song, M., Zhao, X., Haihong, E. and Ou, Z., (2016), "Statistic-based CRM approach via time series segmenting RFM on large scale data", *Proceedings of the 9th International Conference on Utility and Cloud Computing*, December, pp. 282-291.confproc

Stewart, T.J. (1992), "A critical survey on the status of multiple criteria decision making theory and practice", *Omega*, Vol. 20 Nos 5–6, pp. 569-586.

Tansley, S. and Tolle, K.M. (2009), *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Vol. 1, Hey, A.J. (Ed.), Microsoft Research, Seattle, Redmond, WA.

Tiwari, A., Gupta, R.K. and Agrawal, D.P. (2010), "A survey on frequent pattern mining: current status and challenging issues", *Information Technology Journal*, Vol. 9 No. 7, pp. 1278-1293.

Topi, H. and Tucker, A. (Eds) (2014), *Computing Handbook: Information Systems and Information Technology*, Vol. 2, CRC Press, Florida.

Tsai, C.Y. and Chiu, C.C. (2004), "A purchase-based market segmentation methodology", *Expert Systems with Applications*, Vol. 27 No. 2, pp. 265-276.

Vohra, R., Pahareeya, J., Hussain, A., Ghali, F. and Lui, A. (2020), "Using self organizing maps and K means clustering based on RFM model for customer segmentation in the online retail business", *International Conference on Intelligent Computing*, October, Springer, Cham, pp. 484-497.

Wang, C., Li, X., Zhou, X., Wang, A. and Nedjah, N. (2016), "Soft computing in big data intelligent transportation systems", *Applied Soft Computing*, Vol. 38, pp. 1099-1108.

Wedel, M. and Kamakura, W.A. (2000), "The historical development of the market segmentation concept", in *Market Segmentation*, Springer, Boston, Massachusetts, pp. 3-6.

Wilson, J., Chaudhury, S. and Lall, B., (2018), "Clustering short temporal behaviour sequences for customer segmentation using LDA", *Expert Systems*, Vol. 35 No. 3, 12250.

Wong, J.Y., Chen, H.J., Chung, P.H. and Kao, N.C. (2006), "Identifying valuable travelers and their next foreign destination by the application of data mining techniques", *Asia Pacific Journal of Tourism Research*, Vol. 11 No. 4, pp. 355-373.

Woo, J.Y., Bae, S.M. and Park, S.C. (2005), "Visualization method for customer targeting using customer map", *Expert Systems with Applications*, Vol. 28 No. 4, pp. 763-772.

Worthington, S. and Welch, P. (2011), "Banking without the banks", *International Journal of Bank Marketing*, Vol. 29 No. 2, pp. 190-201.

Wei, J.T., Lin, S.Y. and Wu, H.H. (2010), "A review of the application of RFM model", *African Journal of Business Management*, Vol. 4 No. 19, pp. 4199-4206.

Yan, B. and Chen, G., (2011), "AppJoy: personalized mobile application discovery", *Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services*, June, pp. 113-126.confproc

Yao, Z., Sarlin, P., Eklund, T. and Back, B. (2014), "Combining visual customer segmentation and response modeling", *Neural Computing and Applications*, Vol. 25 No. 1, pp. 123-134.

Yin, F., Li, C. and Liu, C. (2019), "Study on customer segmentation intelligent model of air cargo", *IOP Conference Series: Earth and Environmental Science*, April, IOP Publishing, Vol. 252 No. 5, p. 052037.

Zhang, M., Zhang, Z. and Qiu, S. (2018), "A customer segmentation model based on affinity propagation algorithm and improved genetic K-means algorithm", *International Conference on Intelligent Information Processing*, October, Springer, Cham, pp. 321-327.

Zhang, Q., Chi, Y., Han, Y., Xu, L., Cheng, C., Zhang, H., Zhang, T., Wu, Y. and Cheng, X. (2021), "An evaluation model of user lifetime value based on improved RFM and AHP method", in *Signal and Information Processing, Networking and Computers*, Springer, Singapore, pp. 991-999.

Zhang, S., Wei, G., Gao, H., Wei, C. and Wei, Y. (2019), "EDAS method for multiple criteria group decision making with picture fuzzy information and its application to green suppliers selections", *Technological and Economic Development of Economy*, Vol. 25 No. 6, pp. 1123-1138.

Zolfani, S.H., Mosharafiandehkordi, S. and Kutut, V. (2019), "A pre-planning for hotel locating according to the sustainability perspective based on BWM-WASPAS approach", *International Journal of Strategic Property Management*, Vol. 23 No. 6, pp. 405-419.

Corresponding author
Madjid Tavana can be contacted at: tavana@lasalle.edu