DOI 10.47176/sjamao.3.1.1

Case Study

SRPH

SJAMAO, 2021; 3(1): 1-8

Customers Grouping Using Data Mining Techniques in the Food Distribution Industry (A Case Study)

Ramin Taghi Livari[1*] and Navid Zarrin Ghalam[2]

[1]Master of Science, Industry Engineering, Islamic Azad University, South Tehran Branch, Thran, Iran.
[2]Master of Science, Industry Engineering, Sharif University of Technology, Thran, Iran.

***Corresponding Author:**
raminlivari@yahoo.com

Received: 15 December, 2020
Accepted: 15 January, 2021
Published: 30 January, 2021

ABSTRACT

Significant data development has required organizations to use a tool to understand the relationships between data and make various appropriate decisions based on the information obtained. Customer segmentation and analysis of their behavior in the manufacturing and distribution industries according to the purposefulness of marketing activities and effective communication and with customers has a particular importance. Customer segmentation using data mining techniques is mainly based on the variables of recency purchase (R), frequency of purchase (F) and monetary value of purchase (M) in RFM model. In this article, using the mentioned variables, twelve customer groups related to the BTB (business to business) of a food production company, are grouped. The grouping in this study is evaluated based on the K-means algorithm and the Davies-Bouldin index. As a result, customer grouping is divided into three groups and, finally the CLV (customer lifetime value) of each cluster is calculated, and appropriate marketing strategies for each cluster have been proposed.

Keywords: Customer relationship management, Data mining, Customer segmentation, RFM analysis, Customer lifetime value

Introduction

Recognizing different groups of customers and establishing effective communication with them in a way that can guarantee the economic benefits of the organization in the future is a very important issue in today's business. Attracting profitable customers as well as retaining valuable old customers are both important. One of the ways to gain knowledge and understanding about customers is to group them and study the characteristics of each group. Considering the intense competition in the market and various options of products and services in front of customers, analyzing customer behavior and choosing the appropriate marketing method based on these analyzes, are considered to be a very important factors for the survival of firms. On the other hand, knowing the correct behavior of customers is one of the most important aspects of customer relationship management. With the rapid development of information technology in various parts of the business and collecting large volumes of customer data, accurate knowledge of the customer, understanding the behavioral pattern of their purchases, adopting marketing strategies and responding to customer needs, has become more difficult. Therefore, it is felt more than ever the need for customer analysis methods based on their bulk data or data mining, the development of appropriate marketing approaches, and communication with customers to satisfy and retain them. Using the customer segmentation approach, one of the data mining techniques, in order to divide heterogeneous customers into homogeneous groups with similar purchasing behavior, will help to better understand customer behavior, an as a result, developing an appropriate marketing strategy and effective communication with each group of customers will be useful in gaining their satisfaction and ultimately the survival of these organizations [1]. In the current and complex business environment, which is constantly growing, customer grouping increases loyalty and improves customer relationships in the long run and ultimately increases profitability for organizations [2]. Data mining helps to identify and track the behavior of customers and their behavioral pattern during the life of their exchanges with the company, which ultimately

[Downloaded from sjamao.srpub.org on 2026-11-14]

[DOI: 10.47176/sjamao.3.1.1]

Publisher: Scientific Research Publishing House (SRPH), Shirvan, Iran
http://srpub.org
Email: sjamao@srpub.org

leads to better customer service, more effective sales, different distribution and marketing strategies, etc. [3]. In this article, we examine the customers of B.T.B of studied company, and their grouping have been studied using the RFM model.

Statement of the problem

Many manufacturing and distribution companies have a huge database of customer information and their purchase history, including customer details, frequency of purchase, amount of money exchanged, and date of customer purchase. It has become clear to the owners of these large businesses that they can use this data and information in all aspects of customer relationship management, including customer identification, acquisition, retention and promotion. One of the most important approaches to customer recognition in organizations that deal with a large number of customers and related data, is the segmentation approach. One of the achievements of this approach is to create a proper interaction with customers in which the product and services provided, have the desired quality and attract their satisfaction [4]. Placing customers with the same number of functional variables in a cluster and the maximum distance of clusters from each other is a measure of clustering quality. One of the models in customer value segmentation and analysis is the segmentation model called RFM model, which was proposed by Hughes in 1994. In this model, three variables of recency of purchase (R), frequency of purchase (F), monetary value of purchase (M), have been proposed as functional variables of customers for their segmentation [5]. Therefore, in this research, the customers are clustered using the k-means algorithm by RFM model. The Davies-Bouldin method was used to evaluate clustering, and finally, by calculating the CLV of each cluster, the customer lifetime value of each cluster is calculated.

Theoretical literature review

Customer relationship management begins with customer recognition, which is the group targeting of people who are profitable for the business. To achieve this, analyzing the behavior of past and present customers will help to understand the characteristics of the organization's customers. Using the segmentation approach and based on the data they have collected from their customers' purchases in the past, organizations divide their customers into groups and categories with similar needs and characteristics in terms of performance variables, so that in the next steps of customer relationship management, develop appropriate strategies to retain and attract customers for each group of customers. Segmentation is one of the most important categories in achieving customer-centric marketing (Battle). Segmentation is one of the data mining techniques and works in such a way that the most similarity occurs within clusters and the most differentiation occurs between clusters [6]. The process of segmenting customers is formed of three main parts of segmentation variables, segmentation methods and segmentation validity. In the following, the application and how to use these variables in this research is discussed.

Segmentation variables: Customer grouping should be done based on the variables that are measured per customer. Variables such as time of purchase, type of customer, frequency of purchase, amount of purchase, etc. are examples of segmentation variables. As mentioned, the RFM model consists of three variables of recency purchase (R) which indicates the time interval of the last purchase of the customer so far and frequency of purchase (F) which indicates the repeat of the customer purchase in a certain period and monetary purchase value (M) which indicates the money value exchanged with the organization is used as effective variables in customer classification. The RFM model is used in customer segmentation to determine the optimal marketing policies. Also, the results of this model are a common tool for the development of marketing strategies. Other application of this model can be used to calculate the value of customer life (CLV).

Segmentation methods: RFM is a descriptive model and there is not a possibility to predict customer behavior in this model. Therefore, each of the segmentation variables is first retrieved only from that aspect of the variable and other variables have no role in its segmentation. Methods and algorithms in the field of clustering are introduced to two categories of hierarchical methods and software. Among the clustering algorithms, the k-means algorithm has received more attention from the group of software methods due to its accuracy and high speed. In this method, which is an iterative algorithm, by assigning k-centers of the cluster, the data is assigned to each cluster with the aim of minimizing the distance of each observation from the center of the cluster. There is a lot of data in a cluster and a lot of differences in the data of adjacent clusters.

Validation of clusters: The number of clusters (k) is an input to the clustering method and the performance of the clustering method depends on the number of optimal clusters that must be specified. For this purpose, the relevant algorithm is executed for the number of different clusters and the number of optimal clusters is determined by indices such as Davis-Boldin index. This index calculates the average similarity of the data of each cluster and to evaluate the lack of similarity between clusters. The lower the number obtained in this index, the better the clustering quality.

[Downloaded from sjamao.srpub.org on 2026-11-14]

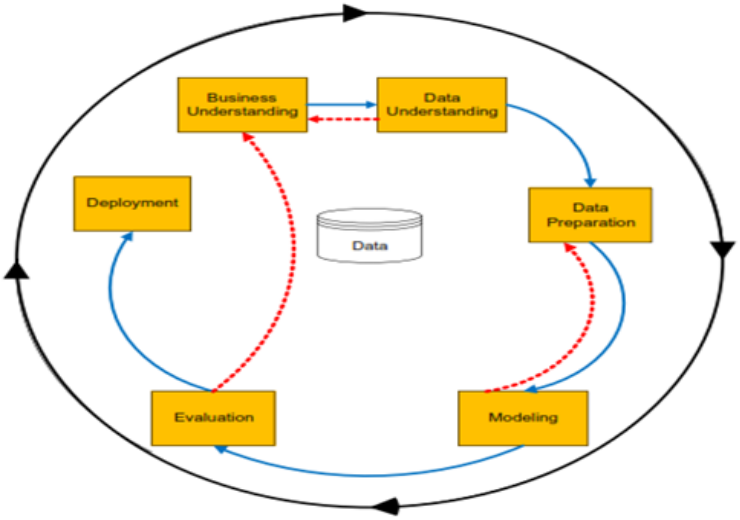
[DOI: 10.47176/sjamao.3.1.1]

Empirical literature review

Anida et al. used the RFM model to analyze customer behavior with the k-means algorithm. The optimal number of clusters in this study was obtained from the Silhouette method [7]. Wu et al. [8] used the RFM model and the K-means clustering method to analyze the value of customers of an industrial equipment manufacturing company. The characteristics of customers were analyzed in the form of clusters and using the value analysis of the customer's lifetime, and suggestions were made to use a suitable promotion program for different segments of customers [8]. Using the RFM model, Monalisa et al. [9] grouped customers with the fuzzy C-means algorithm for LWC. Finally, they calculated the customer lifetime value for the clusters [9]. Migouez said customer grouping is a key to develop and maintain a corporate customer relationship, and developing a customer relationship leads to increased sales. In this paper, a method for market grouping for retail is proposed, which is based on customer lifetime on trading volume, and customer grouping is based on their purchase history [10]. Rabiee combined two methods of RFM and classification to present a model for calculating the value of customer lifetime. Customers calculated and used CLV in C5 algorithm [11]. Mohammadi et al. [12] have used three methods of RFM, FHP and K-means together. They used FAHP technique to weight each variable and finally divided the customers into 5 clusters [12]. In a study presented by Christy et al. [13] customers were grouped based on RFM technique and customer analysis with K-means and fuzzy C-means algorithms [13]. Qadadeh and Abdolah [14] clustered insurance company customers using RFM model and self-made SOM and K-means algorithms. Nguyen et al. [15] studied on the relationship between customer behavior and order in the field of marketing in the online retail company and introduce various marketing tools to increase the level of service to customers [15]. Aran and Kumar, clustered customers using fuzzy and SOM [16]. Shokouhyar et al. [17] used RFM analysis to segment the after-sales service of the car company. They used KANO and SERVQUAL models as tools to measure customer satisfaction [17].

Methodology

In this study, twelve groups of business customers of BTB business that each group of customers has a similar consumer market (for example, customers related to hospitals, universities, municipalities, etc.) belonging to one of the food industry companies were segmented based on the RFM model and based on the k-means algorithm. In this study, the steps of data mining and data analysis in order to discover knowledge of them have been performed by the standard CRISP process. This process includes the steps of understanding the business environment, data selection, data preparation, modeling, model evaluation and development of results (Figure 1). The steps of this method are described separately below.



Step 1. Understanding the business environment.

Considering that identifying and retaining customers has special importance for any organization, in this study we try to identify and group the customers of the study organization. Customer definition: Category Classification of customers into different groups is one of the basic issues of customer relationship management (CRM) that should not be managed in the same way. Customer definition is different in different situations. In the context of "business to business" or BTB, the customers of the organization are an institution (Butel). In this study, due to the importance

[Downloaded from sjamao.srpub.org on 2026-11-14]

[DOI: 10.47176/sjamao.3.1.1]

of knowing specific customers, the grouping of BTB customers has been discussed.

Step 2. Select the data collection

At this step, the necessary information for conducting the research mentioned in the previous steps, was collected. It is need to collect the following information about customers to run the RFM model.
Recency of purchase: Indicates the number of days since the last purchase of the customer; the lower the value of the index, the higher the index value.
Frequency of purchase: Indicates the number of times the customer buys; the higher the index, the higher the index value.

Step 3. Preparation of data

At this step, preparation and preprocessing is done to facilitate the discovery of the hidden knowledge contained in the data. For this purpose, first incomplete and incorrect information and data or invalid and inaccurate values are removed. The data is converted into a format that can be used in the software (RapidMiner).

Step 4. Measure the Segmentation Variables

As mentioned earlier, in the current RFM model, three variables of recency of purchase, frequency of purchase, and monetary value of purchase were used to segment the customers. In the present study, three variables were collected for each customer in a one-year period. Table 1 presents the type of variable collected.

Table 1
Type of collected variables.

Row	Name of variable	Description	Type of variable	
1	ID	Customer code	Integer	Numerical
2	Monetary	Money exchange value	Numeric	Numerical continuously
3	Date	Daily purchase date	Date	Date
4	Recency	Exchange recency	Date	Date
5	Frequency	Exchange frequency	Integer	Numerical

Step 5. Customer Clustering With K-Means Algorithm

The results of applying clustering algorithms to a data collection can be very different depending on the choice of algorithm parameters. The purpose of cluster validation is to find clusters that best fit the data. Two basic measurement criteria proposed for evaluating and selecting the optimal clusters.
Density 10: Data belonging to a cluster should be as close to each other as possible. A common criterion for determining data density is data variance.
Separation 11: The clusters themselves must be sufficiently separated from each other. Three ways are used to measure the degree of separation of clusters: the distance between the nearest data from two clusters, the distance between the farthest data from two clusters, the distance between the centers of clusters.
In this section, first the selection of the number of optimal clusters were evaluated by Davies-Bouldin evaluation index. As a result, with obtaining the minimum index in this evaluation, it was determined that the optimal number of clusters is number 3 and therefore customers were clustered using the K-means algorithm.

Step 6. Determine the Optimal Number of Clusters by Davis-Boldin Index

To determine the optimal number of clusters, the Davis Boldin index was used, the formula of that is as follows:

$$DB = \frac{1}{n}\sum_{\substack{i-1 \\ i\neq j}}^{n} \max\left\{\frac{S_n(Q_i)+S(Q_j)}{S(Q_i.Q_j)}\right\} \quad (1)$$

Where n is the number of clusters and S_n is the average distance between each record and the center of the cluster, and $S(Q_i.Q_j)$ represents the distance between the centers of the clusters. The lowest value in this index indicates the optimal number of cluster selection. The results are presented in Table 2 and the corresponding diagram is presented in figure 2.

Table 2
Davis-Boldin index results.

Index	2	3	4	5	6
DB	0.621	0.495	0.586	0.595	0.501

[Downloaded from sjamao.srpub.org on 2026-11-14]

[DOI: 10.47176/sjamao.3.1.1]

Figure 2. Davis-Boldin index chart.

Given that the minimum value obtained is the optimal number of clusters, then the optimal value for the Davis-Boldin index is number 3. As a result, customers are grouped into 3 clusters.

Step 7. Compare and Evaluate the Results

The results of customer clustering are presented in Table 3.

Table 3
Customer clustering results.

No of records	**2**	**5**	**5**
Cluster	Cluster_0	Cluster_1	Cluster_2
Rial purchase value	1889969500	328290299	887774385
Frequency of purchase	18	10	14
Last of purchase	27	104	46

It can be observed that customers are classified into three clusters. customers of cluster zero are in the valuable customer cluster, customers of cluster one, are in the low valuable cluster and, customers of cluster two, are in medium valuable customers in the organization. Table 4 presents the customer placement in clusters.

Table 4
Placement of customers in clusters.

Cluster_2	Hospitals Organization
Cluster_0	Islamic Republic of Iran Army
Cluster_2	Law Enforcement Force
Cluster_2	Islamic Republic of Iran Navy
Cluster_0	Prisons Organization
Cluster_2	Universities
Cluster_1	Iran Khodro
Cluster_1	SAIPA
Cluster_1	School Buffet Organization
Cluster_1	Social Security Organization
Cluster_1	Ministry of Energy
Cluster_2	Municipality

Next, we label the clusters. In this study, we name higher value customers as golden customers, medium value customers as silver clusters, and lower value customers as bronze clusters (Table 5).

Table 5
Labeling clusters

Cluster	**Label**
Cluster 0	Golden customers
Cluster 1	Silver customers
Cluster 2	Bronze customers

[Downloaded from sjamao.srpub.org on 2026-11-14]

[DOI: 10.47176/sjamao.3.1.1]

Step 8. Calculate the customer lifetime value

In marketing, the lifetime value of each customer is equal to predicting the profitability of that customer in future interactions with that customer. The higher the value of the index, the greater the value of that customer to the organization. After clustering customers to obtain the value of the customer life cycle and after calculating the weights of variables (R, F, M) using formula 2, we calculate the CLV for each customer [9, 11].

$clv = w_m * nm + w_r * nr + w_f$*nf (2)

Where w_m, w_r and w_f are respectively, the weights of the monetary value exchanged, the recency of purchase and frequency of the purchase, which has been calculated using the experts' opinion and Expert-Choice software.

w_m= 0.497

w_r=0.225

w_f= 0.278

And Nf, Nr and Nm and normalization of model variables, respectively, which is done by fuzzy scaling method and the results of calculations, can be seen in table 6 and then the results of customer lifetime value are presented in Table 7.

$$N = \frac{x-\mu}{\sqrt{\sigma^2}} \quad (3)$$

Table 6
Normalization results.

Cluster	Normalized R	Normalized F	Normalized M
Cluster_0	0.95	0.91	0.91
Cluster_1	0.218	0.21	0.1
Cluster_2	0.64	0.55	0.515

Table 7
Calculation of customer lifetime value (CLV) for clusters.

Rank	CLV	w_f*Nf	$w_r * Nr$	$w_m * Nm$	Index
١	٠,٩١٩	٠,٢٥٢٩	٠,٢١٣٧	٠,٤٥٢٢	Golden cluster
٣	٠,١٥٧١	٠,٠٥٨٣	٠,٠٤٩٠	٠,٠٤٩٧	Bronze cluster
٢	٠,٥٥٢٨	٠,١٥٢٩	٠,١٤٤	٠,٢٥٥٩	Silver cluster

The higher the CLV value for a cluster, the greater the value of the cluster customers. As expected, the value of the customer life cycle for customers of golden cluster is higher than the other clusters.

Conclusion

As mentioned, the customers who have added more value to the study company are the Army Organization of the Islamic Republic and the Prisons Organization, which are located in a cluster (cluster 0). Their average lifetime value is 0.919 and only 17% of all customers. Development strategies can be used to maintain and loyalty of this group of customers. Product specificity tailored to the type of cluster of customer development strategies and online ordering for customers of the clusters can help improve customer commitment. Iran Khodro and SAIPA organizations and the School Buffet Organization, the Social Security Insurance Organization, and the Ministry of Energy are located in the bronze cluster (cluster 1). This cluster with an average lifetime value of 0.15 are considered low value customers, which indicates that customers who are in the bronze cluster or there is no platform for cooperation or the potential of these organizations to establish a working relationship is low, this group of customers has a negative attraction strategy. The customers of hospitals, law enforcement, Islamic Republic of Iran Navy, universities and municipalities are in the silver cluster (cluster 2). The average lifetime value of customers in this cluster is 0.55 and 41%. Organizations located in this cluster have the ability to be added to the list of golden cluster customers by creating a suitable platform. In the following, using the types of strategies mentioned in the book of customer relationship management by Francis Butel [18], marketing and customer loyalty strategies are proposed in accordance with the facilities of the study company and the customers of the clusters.

One strategy for customer retention is commitment [18]. Commitment strategy has different types that can be used to explain the appropriate strategy for the characteristics of cluster customers. It is possible to maintain the silver and gold customers of the organization by implementing these strategies.

Instrumental commitment: For the customer of the university organization, instrumental commitment can be used for customer loyalty-creating. In this way, by holding a cooking competition with the company's products in the university, this commitment can be created in the customer and the customer can be persuaded to buy more.

Relationship commitment: Increasing work relationship can help increase customer loyalty and retention [18].

[Downloaded from sjamao.srpub.org on 2026-11-14]

[DOI: 10.47176/sjamao.3.1.1]

This strategy can also be used for hospital customers so that with the presence of quality control officials of the production unit to the customer's location, the health of the products offered to the customer would be ensured. Granting a support card also helps to create a relationship commitment in the customer.

Value-based commitment: When the customer feels that the company's values are in line with the customer's values, he increases his loyalty to the company [18]. This strategy is suitable for the customer of the hospital organization, which is located in a silver cluster. The customer can become a valuable customer of the organization by producing and presenting organic products that can help patients' health and accelerate the recovery process. Other strategies for customer retention include communication strategies [18]. A variety of structural constraints are features of this strategy. Creating these constraints makes it more difficult for the customer to break them.

Financial constraints: This strategy is very important for customers in the BTB business because the most public and private organizations periodically allocate the budget for the purchase of supplies, thus this strategy is both a negative strategy and a positive customer retention strategy, so that credit purchasing facilities are granted to valuable customers.

Process constraints: This strategy works for the school buffet customer. Thus, by presenting and offering individual and student products, the customer's purchasing share can be increased.

Multi-product constraints: When the variety of products that the customer needs, is met by the supplier, it becomes more difficult to break these constraints [18]. Therefore, for the customers of the Army of the Islamic Republic and the Prisons Organization, this strategy is effective for maintaining and building loyalty because they are among the loyal customers and have a high volume of purchases.

Strategy for customer exclusion: Suitable cases for customer exclusion include customers who are not profitable for the organization and are not suitable for implementing any strategy [18]. Bronze cluster customers will be targeted by this strategy. With implementing a negative customer retention strategy for the bronze cluster, in such a way that the amount of customers' purchase can be increased for customers with low value by determining the minimum amount of ordering Rials in each purchase, and if the customer cannot have the minimum purchase can be eliminated, those who are loyal to the organization are identified.

Suggestions

- It is suggested to use the other models such as LRFM, etc. in future researches.
- It is suggested to use association rules to understand the relationships between goods purchased by the customers.
- It is suggested that the customer loyalty strategies of the organization be implemented and measured by customer satisfaction measurement tools, such as KANO, etc., and customer feedback be provided.
- Table 8 summarizes the strategies that can be adopted to the customers of each cluster.

Table 8
Strategies adopted to the customers of each cluster.

Type of strategy	Customer	Description
Relationship commitment	Hospitals organization	Sending quality control officials to check the quality of products
Value-based commitment	Prisons organization	Presentation of organic products
Process constraints	School buffet	Single and student products
Restraint of several products	Prisons Organization / Army of the Islamic Republic	Create restraints by offering more products
Financial constraints	Turn customers into loyal customers	Credit shopping
Instrumental commitment	Universities	Holding conferences and competitions
Negative absorption	Second cluster customers	Determine the minimum purchase amount

References

1. Chang EC, Huang MSC, Wu HH. Using k-means method and spectral clustering technique In an outfitters value analysis quality and quantity. 2010; 44(4): 807-815.

2. Khalili-Damghani K, Abdi F, Abolmakarem S. Hybrid soft computing approach based on clustering, rule mining, and decision tree analysis for customer

[Downloaded from sjamao.srpub.org on 2026-11-14]

[DOI: 10.47176/sjamao.3.1.1]

segmentation problem: real case of customer-centric industries. *Appl Soft Comput.* 2018; 73: 816-828.
3. Han J, Pei J, Kamer M. Data mining: concepts and techniques. *Elsevier.* 2011.
4. Wang YX, Loa Y, Wang Y. A fuzzy-based customer clustering approach with hierarchical structure for logistics network optimization.expert systems with application. 2014; 41(2): 521-534.
5. Wei J, Lin S, Weng C. A case study of applying lrfm model in market segmentation of children dental clinic. *Expert Syst App.* 2012; 39: 5529-5533.
6. Jayasree, Saha, Jayanta, Mukherjee. CNAK: Clustering number assisted K-means. 2021; 110: 107625.
7. Anitha P, Malini, Patil M. RFM model for customer purchase behavior using K-means algorithm. Production and hosting by Elsevier B.V on behalf of king Saudi university. *Comput Inform.* 2019.
8. Wu H-H, En-Chi, Lo C-F. Applying rfm model and k-means method in customer value analysis of an outfitter. International Conference on Concurrent Enginnering, Newyork. 2009.
9. Monalisa S, Nadaya P, Novita R. Analysis for customer life time value categorization with RFM Model. *Proc Comput Sci.* 2019; 834-840.
10. Vlmiguies AS, Camanho, Joafalcaocunha. Customer data mining for lifestyle segmentation. *Expert Syst App.* 2012; 39: 9359-9366.
11. Rabiei A, Hamidrastegari. Integration RFM and classification for response modeing based on customer lifetime value. 2015; 36: 1300-1949.
12. Mohammadi, Bijanbidabad, Nourasteh T, Sherafati M. Credit of bank customers (An integrated model of RFM, FAHP and k-means). 2014; 3: 564-571.
13. Christy A, Joy A, Umamakeswari L, Priyatharsini, Neyaa A. RFMr ranking an effective approach to costumer segmentation. *J King Saudi University Comput Inf Sci.* 2018.
14. Qadadeh W, Abdolah S. Customers segmentation in the insurance company (TIC) dataset.procedia comput. *Sci.* 2018; 144: 227-290.
15. Nguyen, Dung H, Deleeuw, Sander, Dullaert, Wout EH. Consumer behavior and order fulfillment in online retailing: A systematic review. *Int J Manag Rev.* 2018; 20(2): 255-276.
16. Arunachalam D, Kumar N. 2018. Benefit-based consumer segmentation and performance evaluation of clustering approaches: an evidence of data driven decision-making. *Expert Syst Appl.* 2018; 111: 11-34.
17. Shokouhyar S, Shokouhyar S, Safari S. *J Retailing Consum Serv.* 2020; 56: 102139.
18. Butel F, Turnbull J. Customer relationship management book. Siteh Publishing, 2nd Edition.

[Downloaded from sjamao.srpub.org on 2026-11-14]

[DOI: 10.47176/sjamao.3.1.1]

SJAMAO

Copyright: © 2021 The Author(s); This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Citation: Taghi Livari R, Zarrin Ghalam N. Customers Grouping Using Data Mining Techniques in the Food Distribution Industry (A Case Study). SJAMAO, 2021; 3(1): 1-8.

https://doi.org/10.47176/sjamao.3.1.1