Clinical Cancer Investigation Journal-ISSN: 2278-0513

Examining Two-stage Modeling for Customer Segmentation Using Intelligent Clustering Techniques

Abstract

Recognizing the value of customers is a key factor in the success of different stores, and it has become more prominent today. Chain Food stores are in contact with different customer groups and considering their limited resources, they need to rank customers based on their values to be able to allocate an appropriate portion of their marketing resources to more valuable customers to earn more. Therefore, we use data mining techniques to sort customers. Much research is done on this issue. In many studies, the RFM model is used to classify customers. This model consists of three indicators of recency, frequency, and monetary value to analyze customers' purchasing behavior and can determine customers' behavioral value. In this paper, a comprehensive method using segmentation models based on RFM, SOM, SODA, VFT, k-means models is provided, and for customer identification, transactional and demographic data has been investigated. Proposed models were implemented in *chain Food stores* and 350 customers were studied. For transaction data, transactions recorded in the store's information center were used and demographic data were also asked from every customer on the phone. The customers were classified using each of the common model and development model, and in the end, these models were evaluated and compared by the Davies-Bouldin index (DBI) and the sum of squares error (SSE). Based on SSE and DBI, the second model showed better performance in this case study.

Keywords: *Chain Food Stores, Customer Segmentation, Customer Value, Data Mining, RFM, Neural Network, SOM,DBI k-means, AHP ,SODA, VFT*

Mahmoud Dehghan Nayeri
Department of industrial management, Faculty of Management and Economics, Tarbiat Modares University, Tehran.Iran
email:mdnayeri@modares.ac.ir

Alireza Nazari
Department of industrial management, Faculty of Management and Economics, Tarbiat Modares University, Tehran.Iran
email:e.alireza7889@yahoo.com

Ahmad Sadegheih
Department of industrial engineering, Yazd University, Yazd, Iran
email:sadegheih@yazd.ac.ir

Masoud sohrabi
Department of industrial management, Faculty of Management and Economics, Tarbiat Modares University, Tehran.Iran
Sohrabimasoud1@gmail.com

1. Introduction

Detecting customer loyalty and customer segmentation are concepts that have become more prominent today as loyal customers have turned out to be the key component of organizational success. Loyal customers voluntarily advertise for the organization, but disloyal customers discredit the brand in public eyes, so, in today's competitive business, loyal customers need to be identified. Since any economic progress needs precise planning of human resources at the organizational level, in the current competitive era, companies need to focus on their key capabilities and resources to gain a competitive advantage and improve their market position. The competitiveness of companies depends on the development of their competencies, and the supply chain is considered a powerful tool for promoting corporate growth and creating competitive advantages. Supply chain operations play a vital role in administrative decisions since they can significantly affect corporate profitability and operational success. Because of the ever-changing business environment

concerning globalization, supply chain issues have entered the list of priorities of senior executives, but there are still managers who only pay attention to the supply chain, to reduce costs or to solve problems. It may be argued that the biggest problem in manufacturing and service organizations, after managing customer relationships, is the proper management of the supply chain and the provision of manufacturing and service requirements. The belief that supply chain management can make companies more responsive to customers .

thus more profitable has led executives to focus on improving supply chain processes. Many organizations and companies have partly recognized the importance of the role and place of supply chain management in their success. In many cases, corporate executives have undertaken projects and studies to improve their supply chain management, both in terms of using information technology tools and in the application of optimization techniques, such as inventory management and control, the use of lean manufacturing, and other such concepts

2-Literature Review

Peng (2016) maintains that a survey of 180 experienced managers that illustrates the modeling of the structural equation of large organizations, led to an analysis of the importance of three-step functioning. Results show that the conditions of domestic supply chain management, especially information technology and human resources, are the main drivers for improving the overall implementation level of supply chain management and corporate performance.

Cai (2013) states that we consider a supply chain in which the producer provides a new product through a third-party logistics provider to a distant market in which a distributor buys and sells them to end customers. The product is perishable, and both its quantity and quality can be reduced during the transportation process. Market demand is random, sensitive to sales prices and also to the product's freshness. We made proper decisions for these three members of the supply chain, including the transport price of the third-party supply providers, the quantity of the producer's transport and wholesale price, and the amount of the distributor's purchase and retail price. We noticed that the presence of a third-party logistics provider in the supply chain has a significant impact on its performance. We provide a motivation scheme for supply chain coordination. It includes two contracts, containing a wholesale market clearance contract between the manufacturer and the distributor, and a wholesale price discount sharing agreement between the manufacturer and the third-party logistics provider. We show that proposed contracts can eliminate the two sources of "double-marginalization" that exist in the three-tier supply chain and force the three parties to synchronize functioning.

He (2013) maintains that knowledge is an important source of competitive advantage so there is much scientific and practical interest in determining factors that contribute to the transfer of knowledge in the supply chain. Power is known as a key factor in the convenient performance of supply chain participation. However, little empirical research has been done on how power affects the acquisition of knowledge and the performance of supply chain partners. The goal of this study is to address this gap by examining the relationship between power, knowledge acquisition, and supply chain performance among supply chain partners in China's main steel producer. A structured survey was used to collect data. To assess the actual and realized power, two independent variables, "availability of alternative options" and "constraints on the use of power," were used respectively. By controlling probabilities, we found that when the supply chain factors have limited alternatives and when the more powerful factor uses less power, the flow of knowledge increases. In addition, we found a positive relationship between knowledge gains and supply chain performance. This study provides a rich source through the empirical generalization of our understanding of how power affects knowledge and practice.

Supply chain management is a key factor in any supply chain and plays a vital role in the survival and continued success of the supply chain in a competitive global market. In having a successful performance, many variables play a role, but the most important factor in today's business is identifying customers' needs and desires through a customer relationship management (CRM) system.

(Optimization, prediction, modeling and simulation, generalization, and decision support), where neural networks can be deployed. Then, they explained the ways neural networks can be used in supply chain management.

In a paper on the use of genetic algorithms in feature selection, Manuel Guerrero et al. (2017) focused on the issue that recognition of the community of an optimization problem is challenging, and includes searching for communities belonging to an assumed network or graph with nodes of the same community creating common features that provide the ability to identify new features or relationships in the network. A large number of methods have been proposed to solve this problem in many research fields such as power systems, biology, sociology, and physics. Many of these optimization methods are used to identify optimal network segments. This paper offers a new generation of genetic algorithms, which includes efficient innovative techniques and modular search operators. In addition, this method provides a flexible way of analyzing the characteristics of a network at different levels of detail based on the needs of the analyzer. The results obtained from networks in different sizes show good performance of new GA compared to other genetic algorithms, including efficient algorithms published in recent years.

In 2017, Rafael Stanley and his colleague used genetic algorithms to design a human-like robot capable of walking effectively. This task is presented as an optimization problem. The target function is the so-called transport cost and the limitations of the problem are limited to the walking stability criterion of the cycle.

In 2017, Ligang Zhou and his colleagues [1]wrote a paper on the application of a genetic algorithm in a decision tree, which predicts the listing of Chinese companies (PLSCLC). It is an important and complex problem for investors in China. There is plenty of listing

information for each company. We propose an improved filtering method for selecting effective features to predict the listing situation of Chinese companies. Concerning the practical concerns of financial analysts about the performance and interpretation of predicted models, models based on C4.5 and C5.0 decision trees are used and several models are thoroughly deployed. To evaluate the strength of models with time, the models are also tested under Windows. Experimental results show the effectiveness of the proposed method and decision tree model C5.0.

3- Introducing the proposed model

In this section, we introduce our research method based on the type of objective and data collection method and the paper's environment and time. The article is descriptive-exploratory and applied based on its objective. The spatial domain of the article is the Chain Food stores Company. For presenting the models in this paper, a library and field study was done. Transaction data collection was performed using customer records in the company's database and computer networks, and demographic data were obtained by calling the customers. The AHP method has been used to measure the weight of RFM variables following expert opinion. This paper is a cross-sectional study since we studied customers for 12 months in 2021

Matlab and SPSS Clementine software were used in this research. The distinctive feature of Clementine is that it processes the data using nodes connected to form a flow. In addition, after the completion of the data mining process, illustrated data is provided to users. Clementine software's graphical interface invites the user to apply his business skills, which leads to stronger prediction models and a shorter time solution.)

4- Statistical population and sampling method

The statistical population of the study is customers of Chain Food stores stores Companies, surveyed in one year in 2021. The data of this research includes 3800 records related to the customers' transactions. After deleting the missing data, this number reached 3600 records. A simple sampling method was used for sampling and Cochran formula (1) was used to determine the sample size. The sample size obtained with this formula is 350 customers.

(1) $n = \frac{NZ^2P(1-P)}{d^2(N-1)+Z^2P(1-P)}$

N = size of the society

n = sample size

Z = normal value of the standard unit, which at 95% confidence level equals 1.96

P = ratio of the attribute in the community. If it is not available, it can be considered as 0.5. In this case, variance reaches its maximum.

d = amount of error allowed

According to the literature review, RFM model indicators, recency (R), frequency (F), and monetary (M) values are used to evaluate customer behavior. In this paper, the three indicators are used to measure customer behavioral value with SODA and VFT ; also, the Self-Organizing Map (SOM) neural network and k-means are used for segmentation in the design development models.

5-common models :

1- Determining the relative weight of RFM

2- WRFM(recency, frequency, and monetary)

3-K-clusters and cluster centers

4-Segmentation with SOM(self-organization map)

6-Development model:

1- Determining the relative weight of RFM based on SODA (Strategic options Development &Analysis) & VFT (Value Focused thinking algorithm)

2- WRFM

3- Determining the optimal K value for each cluster based on DBI&AHP

4- Segmentation with SOM and k-mean algorithm

To evaluate the validity of the model, DBI and the sum of squared errors(SSE) were used.

According to what is stated above and the definition of similarity between two clusters, DBI is defined as:

$$DB = \frac{1}{n_c}\sum_{i=1}^{n_c} R_i \quad (2)$$

$$R_i = \max_{j=1...nc, i\neq j}(R_{ij}), i = 1...n_c \quad (3)$$

This index measures the average similarity of each cluster with the most similar cluster. It can be seen that the smaller the index is, the better the clusters have been produced.

The sum of square errors index has also been used to evaluate and compare the quality of the two segmentation

models. This index is calculated as (Huang and Kachadi, 2013):

(4) $SSE = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \|c_i - o_{ij}\|^2$

Data used in this research is from chain food stores Company. The data describes the transactions carried out by the customers of the company. Due to limitations on asking for customers' demographic data, three of them were examined; nominal characteristics: gender and education; and a numerical characteristic, age. The age range of 350 customers sampled is from 10 years to 70 years, which is transmitted with the formula (5) to zero to one range.

(5) $x_1 = \frac{x_0 - 11}{70}$

Table1. Segmentation of 350 customers based on demographic features in a SOM neural network

Demographic features				**Number of Segments**	
Education	age	Sex (male)	Number	y	**X**
High school	13-70	73	73	0	**0**
Master's	22-54	46	59	2	**0**
High school	10-41	0	16	0	**1**
PhD	26-53	13	13	1	**1**
Master's	69	1	1	2	**1**
PhD	36-49	0	7	0	**2**
Bachelor's	46-68	22	22	2	**2**
Bachelor's	18-50	0	44	0	**3**
Bachelor's	19-45	114	114	2	**3**

Table 2: Segmentation of 350 customers based on demographic variables in a SOM neural network

Demographic features				**Number of Segments**	
Education	Age	Sex (male)	Number	y	**x**
Bachelor's	19-55	132	132	0	**0**
Bachelor's	18-50	0	44	2	**0**
Bachelor's	58-68	4	4	0	**1**
PhD	36-49	0	10	2	**1**
High school	55-70	2	3	0	**2**
PhD	26-53	10	10	1	**2**
Master's	22-40	0	14	2	**2**
High school	10-50	71	87	0	**3**
Master's	24-54	46	46	2	**3**

Table3. Optimal k- value

Optimal k value for each segment	**Number of Segments**	
	Y	**X**
6	0	**0**
4	2	**0**
3	0	**1**
4	0	**2**

5	1	**2**
4	2	**2**
5	0	**3**
4	2	**3**

Table4. Comparison of the proposed comprehensive model with the k-means model using DBI and the SSE

ceritaria	Model 1	Model 2
DBI	**0.067**	**0.060**
SSE	**1.271**	**1.092**

As you can see in Table 4, in the studied place, based on both DBI and the SSE, the proposed comprehensive model is more efficient than the k-means model.

7- Conclusion

Customer segmentation is an important issue in today's competitive business environment. Many studies have examined the application of data mining technology in customer segmentation and observed its effects. In the literature review of this research, several methods were mentioned for customer segmentation in different industries. According to Cutler's views, markets are divided according to demographic, geographic, psychological, and behavioral variables by. Customer segmentation. in this paper are used VFT and SODA in a behavioral segmentation model. That is, customers are divided based on their behavior (number of transactions, amount of transactions, and recency of transactions) and demographic variables (sex, age, and education). Focusing on value-based customer segmentation, allows stores to use their communication and marketing strategies to be linked to the most valuable customers to maximize their profit and income growth. Sometimes in stores, the strategic interests of segmentation are not paid enough attention. By providing a framework, customer segmentation helps the organization to better select its target group and use its limited resources optimally and effectively to provide satisfaction and thus bring about more profitability. In this paper, 350 customers of a chain store food in the range of 10 to 70 years of age were examined. These customers were classified using a descriptive method consisting of two models and each segment was ranked based on its value. Finally, DBI and the SSE index were used to evaluate and compare the models. Evaluation of DBI and SSE shows that the Second model has a higher quality at this location.thus used SODA and SFT and k-means in second model and compared it with common models on the two criteria stated, the proposed comprehensive model appears to be more effective .

References

1. Cai, x, (2013), Fresh-product supply chain management with logistics outsourcing, Omnega, Volume 41, Issue 4, August 2013, Pages 752–765
2. Carvalho, D. R., & Freitas, A. A, (2004)," A hybrid decision tree/genetic algorithm method for data mining", Information Sciences, Vol.163, pp.1-18.
3. (2006), "Genetic algorithms for optimization of predictive ecosystems models based on decision trees and neural networks", Ecological Modelling, Vol.195, pp.1-5.
4. Dehuri, S., Patnaik, S., Ghosh, A., & Mall, R., (2008), "Appl, Gray, J. B., & Fan, G. (2008), "Classification tree analysis using target". Computational Statistics & Data Analysis, Vol.52, pp.1-3.
5. Hsu, P. L., Lai, R., Chiu, C. C., & Hsu, C. I.,"(2003), the hybrid of association rule algorithms and genetic algorithms for tree induction: an example of predicting the student course performance". ExpertSystemswithApplications, Vol.25, pp.1-11.
6. Huang, C. L., Chen, M. C., & Wang, C. J,"(2007), Credit Scoring with a data mining Approach Based

on Support Vector Machines", Expert Systems with Applications, Vol.33, pp.1-3.

7. Kim, M. J., & Han, I., (2003), "The discovery of experts' decision rules from qualitative bankruptcy data using genetic algorithms". Expert Systems with Applications, Vol.25, pp.1-8.
8. Kim, Y. S., & Sohn, S. Y, (2004), "Managing loan customers using misclassification patterns of credit scoring model", Expert Systems with Applications, Vol.26, pp.1-3.
9. Lee, T. S., & Chen, I. F,(2005), "A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines", Expert Systems with Applications, Vol.28, pp.1-8.
10. Lee, T. S., Chiu, C. C., Lu, C. J., & Chen, I. F,(2002), "Credit scoring using the hybrid neural discriminant technique", Expert Systems with Applications, Vol.23, pp.1-8.
11. Lee, T. S., Chiu, C. C., Lu, C. J., & Chen, I. F,(2002), "Credit scoring using the hybrid neural discriminant technique", Expert Systems with Applications, Vol.23, pp.1-8.
12. Liu, H. H., & Ong, C. S,(2008), "Variable selection in clustering for marketing segmentation using genetic algorithms", Expert Systems with Applications, Vol.34, pp.1-6.
13. Martinez-Otzeta, J. M., Sierra, B., Lazkano, E., & Astigarraga, A., (2006), "Classifier hierarchy learning by means of genetic algorithms". Pattern Recognition Letters, Vol.27, pp.1-6.
14. Nanni, L., & Lumini, A, (2009), "An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring", Expert Systems with Applications, Vol.36, pp.1-4.
15. segmenting Mall Customers Data to improve Business into Higher Target using K-Means clustering December 2021
16. Doi:10.1109/ICAC3N53548.2021.9725630, Conference:2021 3rd international conference on, Advances in computing, communication Control and Networking(ICAC3N)
17. Explainable Customer Segmentation Using K-Means Clustering
    December 2021, Riyo Hayat Khan, Dibyo Fabian Dofadar, Md Golam Rabiul Alam
    BRAC University,
    Doi:10.1109/UEMCON53757.2021.9666609,
    Conference:2021 IEEE12th Annual Ubiquitous Computing, Electronics&Mobile Communication, C conference(UEMCON)
18. Customer Segmentation Using K-means Clustering and the adaptive Particle Swarm Optimization algorithm
    Yue Li, Xiaoquam Chu, Dong Tian, Jianying Feng, Weisong Mu,
    https://doi.org/10.1016/j.asoc.2021.107924
19. Approaches to Clustering in Customer Segmentation, Shreya Tripathi.Aditya Bhardwaj,Poovammal E
    International Journal of Engineering &Technology,7(3.12)(2018)802-807
20. Research and Application of Improved Clustering Algorithm in Retail Customer Classification Chu Fang and Haiming Liu Symmetry 2021.13,1789.https:// doi.org/10.3390/sym13101789.