

# Global E-commerce Market Segmentation by Using Fuzzy Clustering



Basar Oztaysi and Mert Kavi

**Abstract** Customer segmentation is essential for marketing, communication, and even operations management activities. E-commerce provides the data required for novel perspectives to customer segmentation. In this study, we focus on customer segmentation based on purchase variety. To this end, first, the data is preprocessed, and the optimal customer number is detected. Then the fuzzy c-means algorithm is applied, and the segments are formed.

**Keywords** Fuzzy segmentation · Fuzzy c-means · E-commerce

## Introduction

The process of buying, selling, transferring, or exchanging products, services, and/or information via computer networks is defined as E-Commerce [1]. Starting from the invention of electronic fund transfer, and the development of internet e-commerce applications flourished, and various business models have emerged [2]. A standard classification of E-commerce is by the nature of the transactions or the relationship among the participants. The major types can be listed as business-to-business, business-to-consumer, consumers-to-business. As mobile technologies emerge and adapt to social media increases, new e-commerce models such as mobile commerce and social-commerce have become popular. E-commerce can take different forms in a company, depending on the technology.

Customer segmentation is the practice of separating customers into different, meaningful, and identical subsets according to various attributes. For marketing efforts, it is generally used for gathering insights about the customers and provide customized actions according to their characteristics. To achieve this, a company

---

B. Oztaysi (✉)

Industrial Engineering Department, Management Faculty, Istanbul Technical University, Istanbul, Turkey

e-mail: [oztaysib@itu.edu.tr](mailto:oztaysib@itu.edu.tr)

M. Kavi

Modanisa.Com Head Office, Altunizade, Kuşbakişı Cd. No: 27/1, 34662 İstanbul, Turkey

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

233

F. Calisir (ed.), *Industrial Engineering in the Internet-of-Things World*,

Lecture Notes in Management and Industrial Engineering,

[https://doi.org/10.1007/978-3-030-76724-2\\_18](https://doi.org/10.1007/978-3-030-76724-2_18)

needs to understand the customers' segments and the properties and desires of the segments [3]. Companies apply various forms of segmentation, such as value-based, lifetime period, or demographics based. Also, different sources of data can be included in the process, and segments can be formed based on demographics and characteristic properties, and their needs and loyalty levels [4].

Segmentation can be used in three decision levels, strategic, managerial, and operational level [5]. At the strategic level, segmentation is used to assist the corporate strategic plans and highlight product positions according to diversified ranges of buyers. At the managerial level, it is used to deal with resource assignments and targets setting based on customer segments in order to involve the customer groups into the marketing activities and compose organizational processes considering them.

This study aims to summarize potential segmentation perspectives and apply fuzzy clustering with a real-world case study. Fuzzy sets proposed by Zadeh [34] provide tools and operations for handling imprecision and vagueness in real-world problems. Fuzzy sets have been used for many other real-world problems such as, company selection Estralla et al. [29] technology selection [6], multi-criteria decision making Yatsalo et al. [33], public transportation [7], economic analysis [8], and risk management [9]. The case study is from Modanisa.com, which is one of the top e-commerce websites which focus on selling textile products globally.

The remaining of the paper is as follows, in the second section, a brief literature review is provided, then in the third section, fuzzy clustering is introduced. Section four presents the real-world application steps and the results of the study. Finally, the conclusion is given in the last section.

## Literature Review

The literature provides various approaches and methods for customer segmentation, such as demographic-based segmentation, value-based segmentation, propensity based segmentation, life period segmentation. Some of the procedures use heuristics and expert opinions for segmentation, some of them use basic arithmetic operations. The most popular analytical tool for segmentation is cluster analysis. It is a data mining technique that assigns data elements to unknown groups with considering high similarity [10]. The literature also provides various clustering techniques that can be grouped as partitional or hierarchical algorithms. The partitional methods focus on grouping the data elements into a predefined number of clusters. On the other hand, hierarchical methods take advantage of enabling the optimization of a criterion considering the compatibilities of objects within clusters or incompatibilities between clusters.

In recent years various perspectives on market segmentation have been investigated in the literature. Ahani et al. [11] focus on integrating social media data to segmentation efforts to propose a market segmentation model and choice prediction model for SPA hotel market. Liu et al. [12] develop a market segmentation approach by integrating preference analysis and multi-criteria decision-making methods. In the

proposed approach, additive value function and pairwise comparison matrices are used to gather preference data. Then this data is used with the hierarchical clustering method to form the segments. Lim et al. [13] develop a market segmentation approach by using a Bayesian spatial profile regression. By this approach, both spatial autocorrelations present in warehouse rents and multicollinearity among the known rent price determinants are handled for segmentation. Diaz Perez and Bethencourt-Cejas [14] focus on the segmentation of tourists that visit a location and propose using the Chi-square Automatic Interaction Detection method, which is a multivariate analysis technique. The results reveal that the proposed approach is superior to the traditional methods used in the tourism domain. Qin et al. [15] focus on market segmentation problem for demand-side platforms. In their study, the researchers model the optimal market segmentation granularity as an optimization problem and form a mathematical programming model to find the optimum granularity. Huerta-Munoz et al. [16] addresses a customer segmentation problem from a beverage distribution firm. Under a given requirement, the researchers try to form customer clusters where similar customers are involved in. The researchers propose a mathematical attribute formulation and use a greedy heuristic that iteratively destroys and reconstructs customer segmentations. Hong [17] proposes a novel segmentation method by employing the Taguchi method. The Taguchi method is used as a tool to select the initial seeds. The author compares the results with the results of a Self-Organizing Map and shows that the proposed method is superior. Oztaysi and Cevik Onar and Oztaysi [18] propose a user segmentation approach by using twitter data. The author uses the data from social media and uses a fuzzy *c*-means algorithm to segment the users. Oner [19] propose a two steps segmentation approach, in the first step, they use hesitant fuzzy sets to segment the retail locations, and then in the second step, they use customer segmentation by adding the results of the first step into the analysis. Murray et al. [20] deal with a real-world customer segmentation problem where the existing descriptive variables are not suitable for defining the similarities between customers. The authors employ data mining techniques and identify behavior patterns in historical data that involve noise. The authors claim that the proposed results are suitable for strategic decision-making. In a recent study [21] propose a multi-criteria decision-making model for determining the approach for customer segmentation. The authors use Neutrosophic sets to select the most propose approach. Oztaysi et al. [31] propose a segmentation approach based on customer location data. Dogan et al. [22] focus on customer segmentation based on indoor customer paths. Oner and Oztaysi [19] focus on retailer clustering based on hesitant fuzzy sets. Oztaysi [23] use fuzzy *c*-medoids clustering for gender prediction.

## Methodology

In the domain of segmentation, the clustering procedure is used to define subgroups of customers who have common properties. Each customer is defined by a data point in a multi-dimensional space where each dimension represents different properties.

There are various techniques in the literature which convert input data into clusters. Chen et al. [24]. From one perspective, these techniques can be grouped as crisp and fuzzy. The main difference between the two groups is the definition of membership of an element to a cluster. Crisp clustering assigns a data element into a single specific cluster, while fuzzy clustering algorithms assign a data element to diversified clusters simultaneously with a membership degree [25]. In the literature, fuzzy clustering is used by using two points of view, either considering uncertain data or considering crisp data with uncertain clusters [26]. One of the most commonly adopted techniques for fuzzy clustering is the fuzzy  $c$ -means algorithm, in which data elements are assigned to a predefined number of clusters with different membership values [24].

The fuzzy  $c$ -means algorithm is based on similarity or dissimilarity measures, which are extracted from distance measurement such as Euclidean distance [27]. The main definition is the partition matrix, which represents the extracted clusters. A fuzzy partition matrix is defined from [28] with the conditions given in the following:

$$\begin{aligned} \mu_{ik} &\in [0, 1], 1 \leq i \leq c, 1 \leq k \leq N, \\ \sum_{i=1}^c \mu_{ik} &= 1, \quad \leq k \leq N, \\ 0 < \sum_{k=1}^N \mu_{ik} &< N, 1 \leq i \leq c \end{aligned}$$

Equation (1b) defines the sum of each cluster should be equal to 1, and the membership degree should be represented with an interval [11].

Fuzzy  $c$ -means clustering utilizes an objective function and focuses on minimizing the objective function to find the appropriate clusters.

$$J(Z, U, V) = \sum_{i=1}^c \sum_{j=1}^N (\mu_{ij})^m \|z_j - v_i\|^2$$

where  $Z$  is the set of data elements to be clustered,  $U$  shows the fuzzy partition matrix,  $V$  is the vector which indicates the cluster centers. As seen from the given formula,  $N$  represents the number of observations,  $\mu$  denotes the related membership value,  $c$  is the number of appeared clusters, and  $m$  is the parameter called fuzzifier that identifies the fuzziness degree of the final clusters, and fuzzifier parameter can get values greater than 1. When the fuzzifier parameter equals to one, then the clusters are formed with crisp clustering. Besides that,  $z_j - v_i$  denotes the distance between observation  $j$  and the center of cluster  $i$ .

The minimization of the given objective function comprised of a nonlinear optimization problem that can be calculated with a wide range of techniques such as iterative minimization and heuristic approaches such as simulated annealing or genetic algorithms. The steps of fuzzy  $c$ -means (FcM) clustering algorithm is defined as in the following [29]:

1. Initialize  $U = [u_{ij}]$  matrix,  $U^{(0)}$
2. At  $k$ -step: calculate the centers vectors  $V^{(k)} = [v_i]$  with  $U^{(k)}$

$$v_i = \frac{\sum_{j=1}^N \mu_{ij}^m \cdot z_j}{\sum_{j=1}^N \mu_{ij}^m}$$

3. Update  $U^{(k)}, U^{(k+1)}$

$$\mu_{ij} = \frac{1}{\sum_{k=1}^c \left( \frac{\|z_j - v_i\|}{\|z_j - v_k\|} \right)^{\frac{2}{m-1}}}$$

If  $\|U^{(k+1)} - U^{(k)}\| < \delta$  then STOP; otherwise, go to step 2.

The number of clusters must be determined to reach meaningful clustering results. To this end, various clustering tests are accomplished with different cluster numbers. The results of the tests are compared by using the Xie-Beni index [30], and the clustering with the lowest value is selected for segmentation.

$$X(Z; U, V) = \frac{\sum_{i=1}^c \sum_{j=1}^N (\mu_{ij})^m \|z_j - v_i\|^2}{c \cdot \min_{i \neq j} \|v_i - v_j\|^2}$$

## Results

Modanisa.com is one of the leading e-commerce companies located in Turkey, which focuses on selling textile products to more than 150 countries worldwide. The company sells 70.000 products of 650 different brands. Segmenting customers from different countries is very important for marketing activities. While there can be various segmentation perspectives, segmentation based on purchase variety is selected for this study. Purchase variety refers to the purchase activities of the customers based on several categories the purchased, the number of brands they preferred, and the number of different products they purchased.

The first step of the application is data preparation. The source of the data needed for the segmentation analysis is purchase transactions. The first step of the application is the preprocessing step. The purchase transactions include Product Id, customer Id, quantity data, so Category Id and Brand Id values are added to the transaction table by using SQL commands. Then by raw data that will be used for segmentation are formed. A sample data set is given in Table 1.

The next step is outlier detection and normalization. The values in each column are checked for z values, and the values larger than 3 and lower than  $-3$  are excluded from the data set. The z scores of the remaining data are used for fuzzy clustering.

**Table 1** Sample dataset used for segmentation

Customer Id	#Categories	#Brands	#Products
1,000,000,001	23	24	342
1,000,000,002	5	34	167
1,000,000,003	13	45	87
1,000,000,004	8	6	59
1,000,000,005	2	3	26

Fuzzy c-means clustering is used for segmentation. In order to find the exact number of clusters, various different cluster number parameters are selected, and the results are compared by using Xei-Beni index (Table 2).

The lower values of the Xie-Beni index refer to better clustering results; thus, for this study, c value is selected as five, which means in the study, five clusters are formed. The results can be summarized by the centroid table given in Table 3.

The results present very clear and actionable results. As a summary, Cluster 1 is composed of customers with a very wide variety. They buy from various categories, brands, and products. Cluster 2 is loyal to some categories; they buy a variety of products from a very selected number of categories. Cluster 3 is composed of customers who are loyal to brands and products. Customers in Cluster 4 buy from different categories, but the variety of products does not change much. The final cluster, Cluster 5 is composed of customers with focused categories and brands.

**Table 2** Clustering performances by using Xei-Beni index

Number of clusters	#Categories	Number of clusters	#Products
2	$2.23 \times 10^{-06}$	9	$4.41 \times 10^{-05}$
3	$2.01 \times 10^{-06}$	10	$8.13 \times 10^{-05}$
4	$1.43 \times 10^{-06}$	11	$6.27 \times 10^{-05}$
5	$7.13 \times 10^{-07}$	12	$3.52 \times 10^{-05}$
6	$9.78 \times 10^{-07}$	13	$4.95 \times 10^{-05}$
7	$2.23 \times 10^{-06}$	14	$6.73 \times 10^{-05}$
8	$6.58 \times 10^{-06}$	15	$9.33 \times 10^{-05}$

**Table 3** Centroid table

Customer Id	#Categories	#Brands	#Products
1	1.76	1.93	2.02
2	-1,68	0.4	1.78
3	0.36	-2.01	-1.5
4	1.26	0.4	-0.91
5	-0.89	-1.45	0.21

## Conclusion

In this study, a real-world case study is examined, and steps of segmentation by using the fuzzy *c*-means algorithm is given. After preprocessing, the data are normalized, and the outliers are eliminated. A set of *c* parameter is used for different runs of fuzzy *c*-means. Xie-Beni index is used to obtain the best parameter, and the result reveals that five clusters are best for segmentation.

For further studies, other perspectives of clustering, such as value-based clustering or intention based clustering, can be examined. Besides, other fuzzy and crisp clustering methods can be used with the same data, and the results can be compared with the outcome of this study.

## References

1. Turban E, King D, Liang PL, Turban D (2012) *Electronic commerce 2012: managerial and social networks perspectives*. Prentice Hall, Upper Saddle River
2. Sharma V (2012) A Study about management and business issues of E-commerce. *IOSR J Bus Manage* 2(2):1–5. <https://doi.org/10.9790/487X-0220105>
3. Wind YJ, Bell DR (2007) Market segmentation. In Baker M, Hart S (eds) *The marketing book*. Butterworth Heinemann
4. Tsitsis K, Chorianopoulos A (2009) *Data mining techniques in CRM: inside customer segmentation*. Wiley
5. Cravens WD, Piercy NF (2012) *Strategic marketing*. McGraw-Hill Higher Education
6. Dogan O, Oztaysi B (2018) In-store behavioral analytics technology selection using fuzzy decision making. *J Enterp Inf Manag* 31(4)
7. Kaya I, Oztaysi B, Kahraman C (2012) (2016) A two-phased fuzzy multi-criteria selection among public transportation investments for policy-making and risk governance. *Int J Uncertain Fuzziness Knowl Based Syst* 20(supp01):31–48
8. Kahraman C, Cevik Onar S, Oztaysi B (2016) A comparison of wind energy investment alternatives using interval-valued intuitionistic fuzzy benefit/cost analysis. *Sustainability* 8(2):118
9. Behret H, Öztayşi B, Kahraman C (2011) A fuzzy inference system for supply chain risk management. In: Wang Y, Li T (eds) *Practical applications of intelligent systems. Advances in intelligent and soft computing*, vol 124. Springer, Berlin
10. Han J, Kamber M, Pei J (2011) *Data mining: concepts and techniques*, 3rd edn. Morgan Kaufmann Publishers
11. Ahani A, Nilashi M, Ibrahim O, Sanzogni L, Weaven S (2019) Market segmentation and travel choice prediction in Spa hotels through TripAdvisor's online reviews 80:52–77
12. Liu J, Liao X, Huang W, Liao X (2019) Market segmentation: a multiple criteria approach combining preference analysis and segmentation decision. *Omega* 83:1–13
13. Lim H, Yoo HE, Park M (2018) Warehouse rental market segmentation using spatial profile regression. *J Transp Geogr* 73:64–74
14. Diaz-Perez FM, Bethencourt-Cejas M (2016) CHAID algorithm as an appropriate analytical method for tourism market segmentation. *J Destin Mark Manage* 5(3):275–282
15. Qin R, Yuan Y, Wang FY (2017) Exploring the optimal granularity for market segmentation in RTB advertising via computational experiment approach. *Electron Commer Res Appl* 24:68–83
16. Huerta-Munoz DL, Rios-Mercado RZ, Ruiz R (2017) An iterated greedy heuristic for a market segmentation problem with multiple attributes. *Eur J Oper Res* 261(1):75–87

17. Hong CW (2012) Using the Taguchi method for effective market segmentation. *Expert Syst Appl* 39(5):5451–5459
18. Oztaysi B, Cevik Onar S (2013) User segmentation based on twitter data using fuzzy clustering. *Data Min Dyn Soc Netw Fuzzy Syst* 316–333
19. Oner SC, Oztaysi B (2017) An interval valued hesitant fuzzy clustering approach for location clustering and customer segmentation. *Adv Fuzzy Logic Technol* 2017:56–70
20. Murray PW, Agard B, Barajas MA (2017) Market segmentation through data mining: a method to extract behaviors from a noisy data set. *Comput Ind Eng* 109:233–252
21. Kahraman C, Cevik Onar S, Oztaysi B (2019) Customer segmentation method determination using neutrosophic sets. In: International conference on intelligent and fuzzy systems, pp 517–526
22. Dogan O, Oztaysi B, Fernandez-Llatas C (2020) Segmentation of indoor customer paths using intuitionistic fuzzy clustering: process mining visualization. *J Intell Fuzzy Syst* 38(1):675–684
23. O Dogan, B Oztaysi (2019) Gender prediction from classified indoor customer paths by fuzzy C-medoids clustering. In: International conference on intelligent and fuzzy systems, pp 160–169
24. Chen S, Fern A, Todorovic S (2014) Multi-object tracking via constrained sequential label-ing. In: Paper presented at IEEE conference on computer vision and pattern recognition, Colum-bus, OH, 23–28 June
25. Oztaysi B, Isik M (2014) Supplier evaluation us-ing fuzzy clustering. In: Supply chain management under fuzziness, pp 61–79
26. Aliahmadipour L, Torra V, Eslami E (2017) On hesitant fuzzy clustering and clustering of hesitant fuzzy data. In: Torra V, Dahlbom A, Narukawa Y (eds) *Fuzzy sets, rough sets, multisets and clustering. Studies in computational intelligence*, vol 671. Springer, Cham
27. Babuska R (2009) Fuzzy and neural control disc course lecture notes. Retrieved 5 Oct 2012 from [http://www.dcsc.tudelft.nl/~disc\\_fnc/transp/fncontrol.pdf](http://www.dcsc.tudelft.nl/~disc_fnc/transp/fncontrol.pdf)
28. Ruspini EH (1970) Numerical methods for fuzzy clustering. *Inf Sci* 2(3):319–350. [https://doi.org/10.1016/S0020-0255\(70\)80056-1](https://doi.org/10.1016/S0020-0255(70)80056-1)
29. Estrella FJ, Cevik OS, Rodríguez RM, Oztaysi B, Martínez L, Kahraman C (2017) Selecting firms in University technoparks: a hesitant linguistic fuzzy TOPSIS model for heterogeneous contexts. *J Intell Fuzzy Syst* 33(2):1155–1172
30. Xie XL, Beni G (1991) A validity measure for fuzzy clustering. *IEEE Trans Pattern Anal Mach Intell* 13(8) (1991)
31. Öztayşı B, Gokdere U, Simsek EN, Oner CS (2017) A novel approach to segmentation using customer locations data and intelligent techniques. In: *Handbook of research on intelligent techniques and modeling applications in marketing analytics*
32. Oner SC (2018) B Oztaysi (2018) An interval type 2 hesitant fuzzy MCDM approach and a fuzzy c means clustering for retailer clustering. *Soft Comput* 22(15):4971–4987
33. Yatsalo B, Korobov A, Oztaysi B, Kahraman C, Martínez L (2020) A general approach to fuzzy TOPSIS based on the concept of fuzzy multi-criteria acceptability analysis. *J Intell Fuzzy Syst* 38(1):979–995
34. Zadeh LA (1980) Fuzzy sets and information granularity. *Adv Fuzzy Set Theory App* 11:3–18