A COMBINATION OF FORENSIC-BASED INVESTIGATION ALGORITHM AND DENSITY PEAK-BASED FUZZY CLUSTERING FOR CUSTOM SEGMENTATION

Thi Phuong Quyen Nguyen*

The University of Danang - University of Science and Technology, Danang, Vietnam

*Corresponding author: ntpquyen@dut.udn.vn

(Received: March 09, 2023; Revised: April 11, 2023; Accepted: April 21, 2023)

Abstract - Custom segmentation is a process of classifying potential customers based on their mutual features such as shopping habits, consumption trends, and demand to provide an effective marketing campaign for each customer group. Data clustering is one of the most common methods for custom segmentation. This study proposed a novel clustering method that employs density peak-based fuzzy *c*-means (DP-FCM) and forensic-based investigation (FBI) algorithms. The proposed method (denoted as DP-FBI-FCM) aims to provide an effective clustering technique that can exploit the global optimal solution for custom segmentation problems. Besides, the proposed DP-FBI-FCM is used to segment wholesale customer data of a supermarket. As a result, four distinct customer groups are classified. Businesses can implement different strategies in each cluster to retain and attract their customers.

Key words - Clustering method custom segmentation; forensic-based investigation; fuzzy c-means.

1. Introduction

Custom segmentation is a process of classifying potential customers based on their mutual characteristics such as shopping habits, consumption trends, and demand. It can assist businesses in better understanding the behavior of customers and implementing various marketing campaigns for specific markets to increase sales and customer satisfaction [1]. The explosion of information technology, especially the development of Industry Revolution 4.0 is currently changing how businesses approach market segmentation. Digital versions of numerous products are now available. A new set of customer behaviors, identities, and expectations have been produced because of the two-way information flow between customers and providers that technology has enabled. Businesses must modify and adjust their segments to take into account new information on actual customer behavior as it changes over time.

Cluster analysis is one of the most widely used methods for market segmentation. Customers are grouped into relatively homogeneous groups using specific criteria. As a result, customers who are placed in the same group will exhibit greater similarities than those who are placed in other clusters [2]. The clustering approach is a powerful method for performing market segmentation because it can effectively categorize the market and analyze both numerical and categorical data [3]. There have been numerous clustering methods proposed for custom segmentation, each with its pros and cons in various fields, such as travel and tourist [4-6], banking and finance [7, 8], retail market, freight market, food industry, clothing, and fashion [9-11]. The most common clustering methods for custom segmentation are *k*-means [12, 13], fuzzy *c*-means (FCM) [14], self-organizing maps [15], and so on. These algorithms evaluate the characteristics and shopping behavior of customers to establish various computations for the cluster center, distance measure between customers, splitting method, threshold, and number of iterations to determine distinct segments. The FCM-based clustering algorithms employ the fuzzy membership function matrix and termination condition to group the customers. Velmurugan [16] made a comparison of using *k*-means and FCM algorithms in custom segmentation for telecommunications data sets. The comparison result showed that *k*-means algorithm was more prominent in terms of computation time since FCM algorithm took more time for the iteration process and fuzzy computation. However, FCM algorithm gave better and more consistent results in terms of clustering accuracy with different number of iterations, fuzzy values and stopping conditions.

The previous study in clustering still has the following drawbacks: 1) They are unable to handle outlier and noise data, although outliers and noise are presented in all datasets; and 2) clustering performance depends on the initial clustering centers, which might produce undesirable results by trapping local optimal solutions. To overcome these challenges, multiple clustering methods had been proposed. For instance, density-based clustering has more advantages compared to various approaches for controlling noise and outlier data [17]. Especially, Liu et al. proposed a density peak clustering based fuzzy *c*-means (DPFCM), where the cluster centers were adopted based on the density distance between each data instance, to not only handle the noise data and the sensitivity of cluster centers but also improve the clustering accuracy [18]. However, the cluster centers in the DPFCM algorithm were chosen manually from the decision graph-based density distance, which is one of its shortcomings because it is difficult to determine [19]. Other studies combined metaheuristic approaches such as genetic algorithms, ant colony, particle swarm optimization, and so on, with clustering methods to enhance the clustering performance and explore the global optimal solution [20-22].

Based on the aforementioned analysis, this study proposes a novel clustering technique to solve the customer segmentation problem. The proposed clustering method employed DPFCM as the fundamental approach and combines the new metaheuristic, i.e., forensic-based investigation (FBI) [23] to improve the segmentation result of the DPFCM algorithm. FBI is a free-parameter search algorithm inspired by the forensic investigation process of police officers. The FBI algorithm consists of two main stages: investigation and pursuit. The investigation process determines the target location, and the pursuit process aims to catch the target. The combination of FBI and FCM algorithms will make it easier to identify a globally optimal solution for the data clustering problem, which can then be successfully applied to segmenting the customer data of the business organization.

2. Related works

2.1. Fuzzy C-Means Algorithm

FCM is a popular clustering method that uses fuzzy logic in data partitioning [24]. The fuzzy concept allows a data instance to belong to a cluster based on its membership degree from 0 to 1. A dataset $X = \{x_1, x_2, \dots, x_n\}$ contains *n* data instances. FCM algorithm classifies *X* into *c* clusters to minimize the following objective function:

$$J = \sum_{i=1}^{n} \sum_{i=1}^{c} [\mu_{ij}]^m \times d_{ij}(x_i, v_j)^2, \quad (1)$$

Where, μ_{ij} is a fuzzy membership function, $\mu_{ij} \in [0,1]$, $\sum_{j=1}^{c} \mu_{ij} = 1$; $1 \le i \le n$; $1 \le j \le c$; $d_{ij}(x_i, v_j)$ is the distance from data instance x_i to its cluster center v_j, *m* is a membership degree which is usually set as 2.

FCM algorithm minimizes the objective function in Eq. (1) by repeating the updating process of cluster center v_j and membership function μ_{ij} as follows:

$$v_j = \frac{\sum_{i=1}^{n} [u_{ij}]^m x_i}{\sum_{i=1}^{n} [u_{ij}]^m}; \ 1 \le i \le n; \ 1 \le j \le c, \quad (2)$$

$$\mu_{ij}{}^{(t+1)} = \left[\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{ik}} \right)^{\frac{2}{m-1}} \right]^{-1} \quad (3)$$

The updating process is repeated until the stopping condition is met.

2.2. Density Peak Fuzzy C-Means Algorithm

DPFCM algorithm is a combination of density peak clustering (DPC) approach and FCM algorithms. The concept behind the DPC is that the cluster centers are surrounded by neighbors who have lower local densities [25]. The DPC approach determines two values: the local density of each data point *i* and the distance between that point and the point with the highest density [26]. Combined with FCM algorithm, the DPFCM has several advantages such as effectively solving the problem of sensitivity to the initial centers, and improving the clustering accuracy [18]. The procedure of DPFCM is described as follows:

Stage 1: Identify the initial cluster centers.

Step 1: For each data instance *i*, compute the local density ρ_i and distance δ_i as follows [25]:

$$\rho_i = \sum_j \chi^{(dis_{ij} - d_c)} \quad (4)$$

$$\delta_i = \min_{j: \rho_j > \rho_i} (dis_{ij}) \quad (5)$$

Where, dis_{ij} is the distance between instances *i* and *j*, d_c is the truncation distance, $\chi^{(dis_{ij} - d_c)} = 0$ if $dis_{ij} - d_c \ge 0$, otherwise, $\chi^{(dis_{ij} - d_c)} = 1$. Regarding Eq. (2), the distance becomes $\delta_i = \max_j (dis_{ij})$ for the data instances that have the highest density.

Step 2: Determine the distance-based density index φ_i using Eq. (3) and then sort φ_i value in descending order.

$$\varphi_i = \rho_i * \delta_i \quad (6)$$

Step 3: Select *z* instances that have the largest distance-based density then calculate their average value, which is denoted as $\varphi_{average}$.

Step 4: Select *k* cluster centers which are the points that have $\varphi_i > \varphi_{average}$.

Stage 2: Implement FCM algorithm. The cluster centers obtained from stage 1 are used as the initial cluster centers for the FCM algorithm.

2.3. Forensic-Based Investigation Algorithm

The FBI algorithm proposed by Chou and Nguyen [23] was inspired by the forensic investigation process without requiring predefined operating parameters. FBI was designed to identify global solutions to continuous nonlinear functions with high accuracy and less computational cost. The processes of investigation, location, and pursuit of suspects by police officers were inspired by the design of the FBI algorithm. Some key features of the FBI algorithm are listed as follows:

(1) FBI is a free-parameter optimization algorithm.

(2) FBI outperforms well-known and recently developed algorithms by a wide margin.

(3) FBI has a faster computation time.

(4) Its structure consists of two teams that successfully balance exploration and exploitation well to obtain the global optimal solution.

The general idea behind the FBI algorithm is shown in Figure 1 [23].

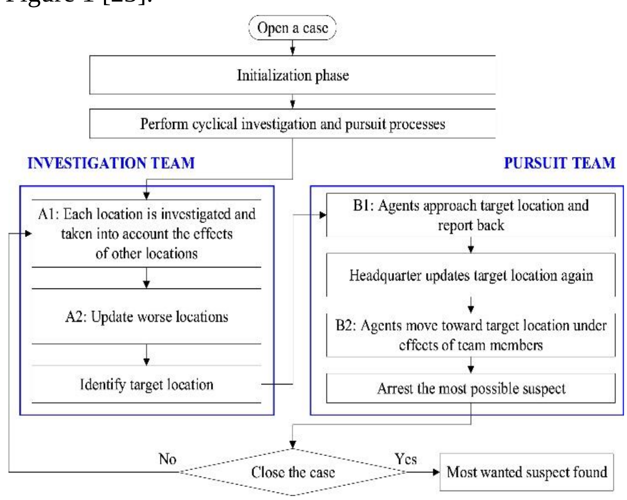


Figure 1. *General procedure of the FBI algorithm [23]*

3. Proposed DP-FBI-FCM algorithm

The proposed DP-FBI-FCM algorithm, which combines DPFCM with FBI algorithms, aims to not only overcome the drawback of sensitivity of initial cluster centers in most of the clustering methods but also can explore the global optimal solution and improve the clustering accuracy. The procedure of DP-FBI-FCM algorithm is described in Figure 2.

The following parts are used to describe clearly how to employ the FBI algorithm based on the procedure in Figure 2. First, the DPC clustering approach is applied to find the cluster centers. The cluster center is represented by a variable X_{ij}. The DPC clustering is implemented multiple times to obtain the population *NP* for the FBI algorithm.

Algorithm 1. DP-GA-PFCM Algorithm	
Input:	k (number of clusters), *MAX_ITER* (maximum of iterations)
Output:	L (set of k cluster label)
Step	
DPC procedure	
1	*Calculate local density ρ_i using Eq. (1)*
2	*Calculate the distance δ_i using Eq. (2)*
3	*Calculate the distance-based density index φ_i using Eq. (3)*
4	*Calculate $\varphi_{average}$ of z instances that have largest distane-based density*
5	*Select top k cluster centers as step 4 of DPFCM algorithm*
6	*Initialize population*
7	Calculate fitness using FCM algorithm
FBI procedure	
8	***while*** *i < MAX_ITER,* ***do***
9	*Implement step A1 of the FBI algorithm to find the possible locations of cluster centers*
10	*Implement step A2 of the FBI algorithm to define and update the target cluster centers*
11	*Implement step B1 of the FBI algorithm to approach the target cluster centers and update them.*
12	*Implement step B2 to explore better cluster centers and increase the accuracy.*
13	***end while***
14	*Select the best candidate*
FCM procedure	
15	*Initialize membership*
16	***while*** *j < MAX_ITER,* ***do***
17	*Calculate typicality matrix*
18	*Update cluster centers*
19	*Update membership*
20	***end while***
21	*Report L as a set of cluster labels*

Figure 2. *The procedure of the DP-FBI-FC algorithm*

Step A1: In step A1 of the FBI algorithm, the cluster centers are updated as follows:

$$X_{A1_{ij}} = X_{ij} + \left((rand - 0.5) * 2\right) * (X_{ij} - \frac{X_{kj}+X_{hj}}{2}), \quad (7)$$

Where, X_{ij} is the initial cluster center or initial location, $X_{A1_{ij}}$ is the new suspect's location at step A1; i, k, h represents the three suspect's locations $(i, k, h = 1, 2, \ldots, NP)$. To update the new location for candidate i, k, and h are randomly selected in the population *NP*; $j = 1, 2, \ldots, D$ with D is the number of dimensions, $\left((rand - 0.5) * 2\right)$ results in the random number in $[-1, 1]$.

The new suspected location is evaluated and compared with the current entry based on the objective function. Herein, the FCM is embedded to calculate the objective function. The cluster centers X_{ij}, and $X_{A1_{ij}}$ are used to implement the FCM algorithm. The objective functions obtained for the current location and the new suspect's location are p_i, and p_{A1_i}, respectively. If p_{A1_i} is better than p_i, the suspect's location X_{ij} is replaced by the new one $X_{A1_{ij}}$. Otherwise, the current entry is kept in the population.

Step A2: The FBI algorithm is continued to update the target location. The probability of a candidate $X_{A1_{ij}}$ is calculated using Eq. (8) as follows:

$$Prob\left(X_{A1_i}\right) = (p_{worst} - p_{A1_i})/(p_{worst} - p_{best}), \quad (8)$$

Where, p_{worst}, and p_{best} are the worst and the best objective values in the population, respectively.

The target location is updated as follows:

$$X_{A2_{ij}} = X_{best} + X_{A1_{dj}} + rand * (X_{A1_{ej}} - X_{A1_{fj}}) \quad (9)$$

Where, X_{best} is the best location corresponding to p_{best}, d, e, f, i are the four candidates in the population. To update the location for candidate i, the locations d, e, and f are selected randomly.

Similar to step A1, the objective values of the updated candidates are calculated p_{A2_i} for step 2. Then these values are compared with p_{A1_i} to determine the best location at step A2.

Step B1: this step describes the process of approaching the best location determined in step A2. The new locations are continuously updated and their corresponding objective values are obtained. If the objective value of the old location is smaller than that of the new location, it is replaced by the new one. The new location is updated as follows:

$$X_{B1_{ij}} = rand * X_{A2_{ij}} + rand * \left(X_{best} - X_{A2_{ij}}\right) \quad (10)$$

Where, X_{best} is the best location exploited in the population from step A2.

Step B2: Move toward the target location to pick up the most possible suspect. From the population, a candidate is randomly selected as an influencer, denoted as $X_{B_{rj}}$, with its corresponding possibility p_{B_r}. The new location is updated using Eq. (11) if $p_{B_r} > p_{B1_i}$. Otherwise, Eq. (12) is used to update the suspect location.

$$X_{B2_{ij}} = X_{B_{rj}} + rand * \left(X_{B_{rj}} - X_{B1_{ij}}\right) + rand * (X_{best} - X_{B_{rj}}) \quad (11)$$

$$X_{B2_{ij}} = X_{B1_{ij}} + rand * \left(X_{B1_{ij}} - X_{B_{rj}}\right) + rand * \left(X_{best} - X_{B1_{ij}}\right) \quad (12)$$

Where, X_{best} is the best location exploited in the population from step B1.

At the end of step B2, the best location and its corresponding possibility are updated as $X_{best} = X_{B2_{best}}$ and $p_{best} = p_{B2_{best}}$ The process is repeated until the termination condition, which is set as the maximum number of iterations, is reached.

4. Result *analysis*

4.1. Dataset and parameter setting

To illustrate the performance of the proposed DP-FBI-FCM algorithm in custom segmentation, wholesale customer data collected in UCI machine learning repository is used for analysis [27]. This dataset presents the yearly spending in monetary units (m.u.) on several product categories. There are eight features with a total of 440 data instances in this dataset. These features describe the annual spending on fresh products, milk products, groceries, frozen foods, detergent and paper products, delicatessen products, retail channels, and regions.

To implement the proposed algorithm, some parameters need to be set up. For the DPC algorithm, the cut-off distance is needed to identify the truncation distance in Eq. (1). Herein, the cut-off distance is defined as 2% which follows the research result of Rodriguez & Liao [25]. Parameters for FBI, FCM, and DPC algorithms follow their original versions. For detail, FBI is a free-parameter approach which only needs to set up the population size and maximum iteration as 80 and 100, respectively. Cut-off distance is set at 2% for DPC while *m* equals to 2 for FCM's setting after some trials and errors.

4.2. Identify the number of clusters (k)

The number of clusters is a predetermined parameter to implement a clustering algorithm. This study employs one of the most common methods to identify the number of clusters, i.e., the elbow method. The FCM algorithm is used in the elbow method by implementing the dataset with different values of *k* and relying on the SSE (sum of squared errors) value to evaluate the clustering results and select the optimal *k*. Figure 3 shows the SSE values of the dataset implemented by the FCM algorithm with *k* selected from 2 to 20. According to the Elbow method, the optimal *k* is selected at 4.

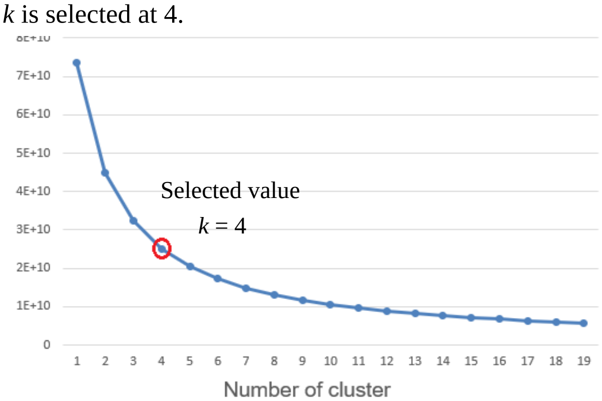


Figure 3. *The scree plot of SSE to select the optimal k*

4.3. Clustering result

The section evaluates the clustering result of the proposed DP-FBI-FCM algorithm. Its result is compared with several benchmark algorithms such as *k*-means, FCM, and DPFCM algorithms. All algorithms are coded in Python and implemented in a Windows 10 computer, Intel Core i5 processor with 16 GB. Each algorithm was run 20 times with its set-up parameter. Then, the average value is presented for comparison.

To evaluate the efficiency of the algorithm, two internal clustering validation indices are employed, i.e., Silhouette (*Si*) and Davies-Bouldin index (DBI) [28]. The pairwise difference between within-cluster distances and between-cluster distances is used to identify the *Si* index. Besides, DBI has calculated the average of all cluster similarities. The higher the *Si* index, the better the clustering outcome. In contrast, the smaller DBI index displays a better clustering result since each cluster has a good degree of separation from the others. Table 1 illustrates the clustering result in terms of *Si* and DBI of the proposed DP-FBI-FCM and other benchmark algorithms. The result shows that the proposed DP-FBI-FCM algorithm outperforms other benchmark algorithms in terms of both *Si* and DBI indices.

Table 1. *The comparison of clustering result*

Index	k-means	FCM	DPFCM	DP-FBI-FCM
Si	0.605	0.538	0.643	0.821
DBI	0.954	1.063	0.906	0.817

4.4. Custom segmentation analysis

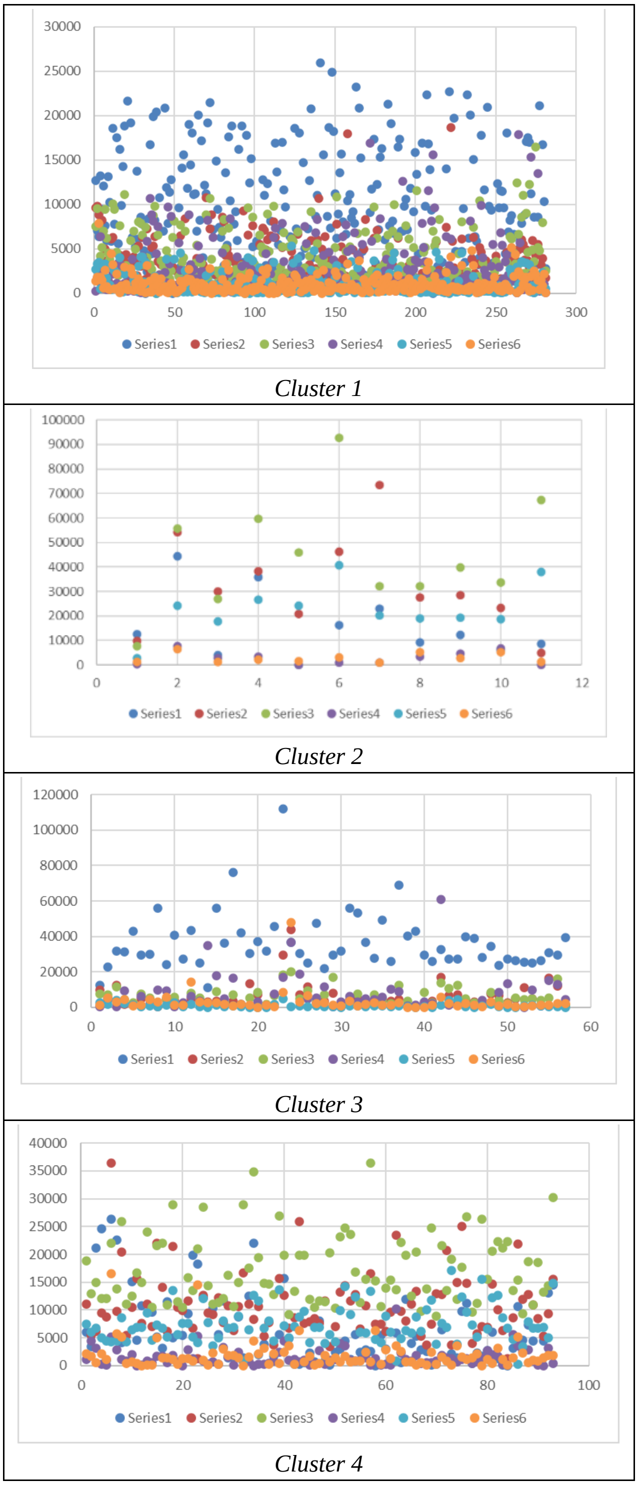
According to the clustering result of the proposed algorithm, the wholesale customer dataset is grouped into 4 clusters. The features of each cluster are presented in Figure 4.

Understanding each customer's behavior is the objective of applying clustering methods to the Wholesale Customers dataset. The proposed clustering algorithm can cluster consumers who exhibit similar behavior. The amount spent on each category of products, as well as the distribution channel in which purchases were made, will be used to describe a customer's behavior. Regarding the Wholesale Customers dataset, the features related to annual spending measured on monetary units, such as annual spending on Fresh products, Milk, Grocery, Frozen products, Detergents and paper, and Delicatessen products, are used to determine the combination of categories that sell together. The features of "Channel" and "Region" are used to identify the customers based on their shopping behaviors by region and sales channel.

Table 2. *Annual spending on product categories of each customer cluster*

	Fresh	Milk	Grocery	Frozen	Detergents _Paper	Delicatessen
Cluster 1	11299.0	3237.4	3348.9	2771.5	5046.5	1133.6
Cluster 2	12421.6	8716.5	17607.1	3646.5	8004.0	1362.1
Cluster 3	21993.6	4754.5	4403.2	4665.3	2149.1	2423.5
Cluster 4	6775.8	5086.7	7576.1	3341.4	1049.1	1344.0

Table 2 shows the customers' segmentation based on the proposed clustering method. Four distinct customer groups are classified. The average spending of each customer cluster on each feature is shown in Table 2. Cluster 1 contains the customers that have low annual spending in the categories of Milk, Grocery, Frozen, and Delicatessen products as well as medium-spending in the remaining categories, which include Fresh and Detergents products. In contrast, the customers in cluster 2 have the highest spending on Milk, Grocery, and Detergents products, while those in cluster 3 consist of customers with the highest spending on Fresh, Frozen, and Delicatessen products. Customers who have low-spending on Fresh and Detergents products as well as medium-spending on Milk, Grocery, and Frozen products are deserving of cluster 4.



Cluster 1

Cluster 2

Cluster 3

Cluster 4

Figure 4. *The features of four distinct clusters*

Besides, Table 3 shows the channel through which each customer cluster bought these products. There are a total of 34 customers in cluster 1. However, most of them are lower spenders in channel 2, which is represented by the Retail channel. Contrarily, 72/79 customers of cluster 2 purchased their goods through the Horeca channel (channel 1). In comparison to clusters 1 and 2, clusters 3 and 4 have significantly more customers-190 and 179, respectively. Customers in cluster 3 are more likely to purchase from sales channel 1 than any other cluster.

Based on the aforementioned analysis of each customer cluster, the distinct characteristic of each customer segmentation is explored. A business organization can be based on these characteristics to have a specific strategy for each segment.

Table 3. *Consumption behavior in terms of sales channel*

	Channel 1	Channel 2	Total
Cluster 1	2	32	34
Cluster 2	72	7	79
Cluster 3	178	12	190
Cluster 4	46	91	137
Total	298	142	440

5. Conclusion

The FBI algorithm is first proposed in this research for market segmentation in a business organization. The FBI algorithm, DPC, and FCM algorithm are then combined to get more accurate segmentation results. To evaluate the performance of the proposed DP-FBI-FCM algorithm, a comparison with the benchmark algorithms, such as *k*-means, FCM, and DPFCM algorithms, was made to precisely assess how well the proposed method performed. The clustering result validated in terms of Silhouette and DBI indices shows that the proposed DP-FBI-FCM algorithm outperforms the benchmark algorithm to cluster the wholesale customer data. Each customer group is then analyzed based on its distinct characteristics to support a specific marketing strategy.

Some directions can be implemented in future research. Regarding the algorithm aspect, improving clustering results is always necessary to obtain a more accurate result. Besides, the researcher can consider the clustering problems for customer segmentation in which customer data contains mix data attributes including numerical and categorical data.

Acknowledgments: This research is funded by the University of Danang - Funds for Science and Technology Development under project number B2020-DN02-83. This support is much appreciated.

REFERENCES

[1] A. Ahani, M. Nilashi, O. Ibrahim, L. Sanzogni, and S. Weaven, "Market segmentation and travel choice prediction in Spa hotels through TripAdvisor's online reviews", *International Journal of Hospitality Management*, vol. 80, pp. 52-77, 2019.

[2] H. Müller and U. Hamm, "Stability of market segmentation with cluster analysis–A methodological approach", *Food Quality and Preference*, vol. 34, pp. 70-78, 2014.

[3] H. Zhou *et al.*, "Market segmentation approach to investigate existing and potential aviation markets", *Transport Policy*, vol. 99, pp. 120-135, 2020.

[4] L. T. Nduna and C. van Zyl, "A benefit segmentation framework for a nature-based tourism destination: the case of Kruger, Panorama and Lowveld areas in Mpumalanga Province", *International Journal*

of Tourism Cities, vol. 6, no. 4, pp. 953-973, 2020.

[5] K. Choi, B. Meng, and T. J. Lee, "An investigation into the segmentation of Japanese traditional 'Ryokan'hotels using selection attributes", *Journal of Vacation Marketing,* vol. 24, no. 4, pp. 324-339, 2018.

[6] L. Nduna and C. Van Zyl, "A benefit segmentation analysis of tourists visiting Mpumalanga", *African Journal of Hospitality, Tourism, and Leisure*, vol. 6, no. 3, pp. 1-22, 2017.

[7] C. Calvo-Porral and J.-P. Lévy-Mangin, "An emotion-based segmentation of bank service customers", *International Journal of Bank Marketing,* vol. 38, no. 7, pp. 1441-1463, 2020.

[8] A. Machauer and S. Morgner, "Segmentation of bank customers by expected benefits and attitudes", *International Journal of Bank Marketing,* vol. 19, no. 1, pp. 6-18, 2001.

[9] C. Fang and H. Liu, "Research and Application of Improved Clustering Algorithm in Retail Customer Classification", *Symmetry*, vol. 13, no. 10, pp. 1789-1793, 2021.

[10] B. Tomassa and E. Rachmawati, "Developing a Marketing Strategy Based on Market Segmentation Through Clustering Method in Food and Beverage Industry in Bandung: Case Study of Bola Ubi Yama", *KnE Social Sciences,* 2020, pp. 1080–1105.

[11] P. Q. Brito, C. Soares, S. Almeida, A. Monte, and M. Byvoet, "Customer segmentation in a large database of an online customized fashion business", *Robotics and Computer-Integrated Manufacturing*, vol. 36, pp. 93-100, 2015.

[12] P. Anitha and M. M. Patil, "RFM model for customer purchase behavior using K-Means algorithm", *Journal of King Saud University-Computer and Information Sciences,* vol. 34, no. 5, pp. 1785-1792, 2022.

[13] M. Aryuni, E. D. Madyatmadja, and E. Miranda, "Customer segmentation in XYZ bank using K-means and K-medoids clustering", in *2018 International conference on information management and technology (ICIMTech)*, 2018, pp. 412-416.

[14] C. Budayan, I. Dikmen, and M. T. Birgonul, "Comparing the performance of traditional cluster analysis, self-organizing maps and fuzzy C-means method for strategic grouping", *Expert Systems with Applications,* vol. 36, no. 9, pp. 11772-11781, 2009.

[15] N. P. P. Yuliari, I. K. G. D. Putra, and N. K. D. Rusjayanti, "Customer segmentation through fuzzy C-means and fuzzy RFM method", *Journal of Theoretical and Applied Information Technology,* vol. 78, no. 3, pp. 380, 2015.

[16] T. Velmurugan, "Performance based analysis between k-Means and Fuzzy C-Means clustering algorithms for connection oriented telecommunication data", *Applied Soft Computing,* vol. 19, pp. 134-146, 2014.

[17] N. Suthar, P. Indr, and P. Vinit, "A technical survey on DBSCAN clustering algorithm", *Int. J. Sci. Eng. Res,* vol. 4, no. 5, pp. 1775-1781, 2013.

[18] X.-y. Liu, J.-c. Fan, and Z.-w. Chen, "Improved fuzzy C-means algorithm based on density peak", *International Journal of Machine Learning and Cybernetics,* vol. 11, pp. 545-552, 2020.

[19] W. Tong, S. Liu, and X.-Z. Gao, "A density-peak-based clustering algorithm of automatically determining the number of clusters", *Neurocomputing,* vol. 458, pp. 655-666, 2021.

[20] R. Ghezelbash, A. Maghsoudi, and E. J. M. Carranza, "Optimization of geochemical anomaly detection using a novel genetic K-means clustering (GKMC) algorithm", *Computers & Geosciences*, vol. 134, pp. 104335, 2020.

[21] L. Zhu, "Safety detection algorithm in sensor network based on ant colony optimization with improved multiple clustering algorithms", *Safety science,* vol. 118, pp. 96-102, 2019.

[22] H. Li, H. He, and Y. Wen, "Dynamic particle swarm optimization and K-means clustering algorithm for image segmentation", *Optik,* vol. 126, no. 24, pp. 4817-4822, 2015.

[23] J.-S. Chou and N.-M. Nguyen, "FBI inspired meta-optimization", *Applied Soft Computing,* vol. 93, pp. 106339, 2020.

[24] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm", *Computers and geosciences,* vol. 10, no. 2-3, pp. 191-203, 1984.

[25] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks", S*cience,* vol. 344, no. 6191, pp. 1492-1496, 2014.

[26] M. Du, S. Ding, X. Xu, and Y. Xue, "Density peaks clustering using geodesic distances", *International Journal of Machine Learning and Cybernetics*, vol. 9, pp. 1335-1349, 2018.

[27] "Wholesale customers Data Set", *UCI machine learning repository*, 2014, [Online]. Availabe: https://archive.ics.uci.edu/ml/datasets/wholesale+customers, [Accessed: Jan. 10, 2023].

[28] Liu, Y., et al. "Understanding of internal clustering validation measures". in *2010 IEEE international conference on data mining*, 2010, pp. 911-916.