Italienisch
ISSN: 0171-4996, Vol. 13, No. 1, 2023, pp 51-59

Fuzzy-Silhouette Model for Automated Determining Loyalty Type User in E-Commerce

I Gusti Agung Gede Arya Kadyanan[1], Luh Gede Astuti[2], Ni Putu Ayu Triana[3]
[1,2,3]Informatic Department of Udayana University, Bali, Indonesia
Email: gungde@unud.ac.id

Abstract

The uniformity of marketing techniques in e-commerce causes failure in its application as a marketing communication strategy. This happens due to the lack of determination of the type of e-commerce user loyalty. Therefore, this study aims to automate the determination of user loyalty types in e-commerce using the Fuzzy-Silhouette model. In its implementation, this model has a range of types in each e-commerce 2-16 types with 4 numerical features that are processed, namely the length of transaction time, the distance between the last transaction and the data processing time, the number of transactions, and the total transaction. For each type of loyalty that has been generated, a standard deviation analysis was carried out to determine the characteristics of that type of loyalty. This research is a quantitative experiment conducted on three e-commerce sites, including UK Retail, Olist Store, and Alnafi Pakistan. The model evaluation results were obtained using the Davied-Bouldin Index (DBI) levels, respectively, of 0.31, 0.36, and 0.24, meaning that the smaller the DBI value obtained, the better the type formation results. So, it can be concluded that the validity of the Fuzzy-Silhouette model for automating the determination of user loyalty types in e-commerce has good results.

Keywords: *Fuzzy-Silhouette, user loyalty types, automation, E-commerce*.

INTRODUCTION

The application of e-commerce is one of the marketing communication strategies used by companies to increase profits in the digital technology era (Irawan, 2022; Jayanti, 2014). E-commerce (electronic commerce) is a business transaction that can be carried out if the two actors are connected via the internet network on a platform supported by electronic devices (Rahiim, 2018). However, the use of e-commerce in improving marketing does not always produce good results because many business people carry out uniform marketing techniques in e-commerce, so the marketing objectives for users are not optimally achieved. As a result of this condition, many businesses have experienced a decline in profits to the point where business operations have ceased, as happened with Nokia (Nasution, 2019; Borhanuddin, 2016). In order to overcome this problem, business people should know the type of loyalty of e-commerce users. The type of user loyalty is a grouping of users using the same similarity values regarding their characteristics such as transaction start date, last transaction date, number of transactions, and total transactions (Alireza & David, 2012). Determining the type of user loyalty in each e-commerce is one of the problems that need to be solved.

Based on the literature review conducted regarding the grouping of e-commerce data, there is research on customer segmentation in e-commerce that is determined using K-Means. This study found that the segmentation results were not optimal because the differences between clusters were not significant (Siagian, Sirait, & Halima, 2021). Other studies also reveal that the use of Fuzzy C-Means is better at producing optimal clusters compared to K-means on dependent data such as e-commerce data. However, the weakness of Fuzzy C-Means is the initialization of the parameters that are determined randomly (Agustina & Prihandoko, 2018).

There are also several studies conducted to overcome the weaknesses of Fuzzy C-Means. Research on the use of optimization methods in the initialization of Fuzzy C-Means centroids so that the results avoid local optimum. The optimization method used is Particle Swarm Optimization (PSO). This method can find the exact initial centroid with a wide search area (Siringoringo & Jamaludin, 2019). There is also research that combines the Gray Wolf algorithm with the Fuzzy C-Means to increase the time effectiveness of the Fuzzy C-Means algorithm (Mohammdian-khoshnoud et al., 2022). There is another study regarding the optimization of the degree of membership of the Fuzzy C Means parameter, which can be done by combining the algorithm with K-Means, which becomes a model called Hybrid OK-Means Fuzzy C-Means (HOFCM) (Perez-Ortega et al., 2022).

Based on this study, there are several methods used to improve Fuzzy C-Means by focusing on increasing the effectiveness of the algorithm's time, optimizing the initial centroid, and optimizing the degree of membership. In this study, the focus is on automating the number of clusters using the Silhouette coefficient function. In this study, the Fuzzy C-Means algorithm will be used as the basic algorithm for the model created and combined with the Silhouette coefficient as an automation function in determining the correct number of types. The proposed model will be referred to as the Fuzzy-Silhouette model. This study uses three e-commerce datasets, including UK Retail, Olist Store, and Alnafi Pakistan, as datasets to evaluate models created using the DBI method. The goal to be achieved in this study is to determine the validity of the Fuzzy-Silhouette model for automating the determination of user loyalty types in e-commerce.

METHOD

The research method used is quantitative experimental research conducted to know the effects caused by treatment in research. This research was conducted in 7 implementation stages, including problem identification, literature study, data collection and preprocessing, implementation of the Fuzzy-Silhouette model, model evaluation, analysis of model results, and presentation of results and drawing of conclusions. An overview of this research design is presented in Figure 1.

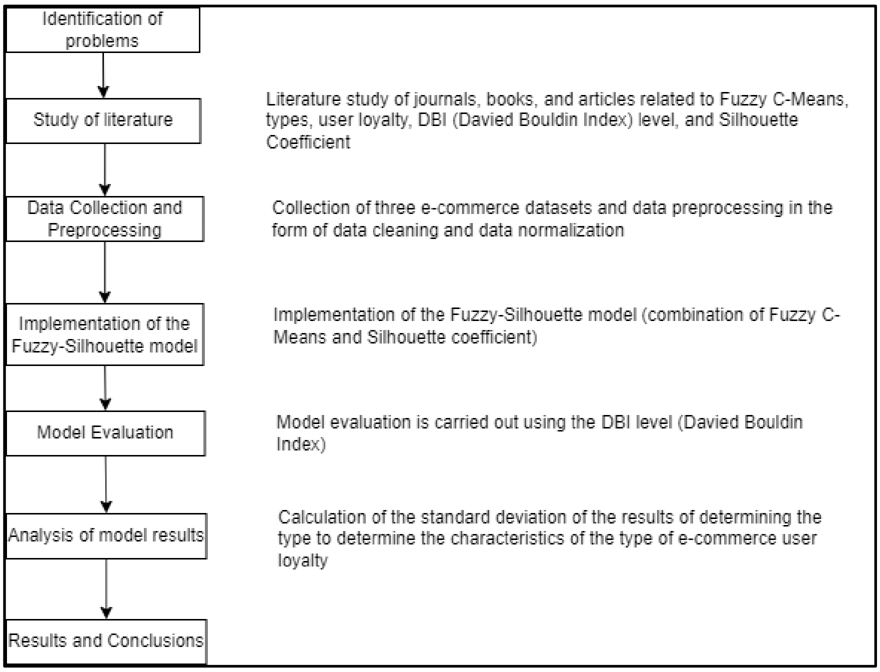

Figure 1. Research Design

The problem identification stage is carried out by identifying the problems faced by business people when using e-commerce as a marketing communication strategy. Identification is carried out to analyze problems to find obstacles faced by business people in implementing e-commerce, especially related to aspects of target marketing to users. That way, researchers are expected to be able to find solutions to address the problem of target marketing to users. After knowing the problems experienced, then proceed to the stage of literature study. The literature study was carried out by looking for a theoretical basis from several previous books, articles, and journals related to Fuzzy-Silhouette, types of user loyalty, e-commerce, and the DBI level (Davied Boludin Index). The purpose of this stage is to strengthen the concept and theoretical basis of the research.

The subjects involved in this research are three e-commerce that has been running including UK Retail, Olist Store, and Alnafi Pakistan. The features used in the created model are taken from the research subjects. The next stage is to collect data from research subjects. The data collected is secondary data obtained from the Kaggle open-source dataset. The amount of data used in each e-commerce is presented in Table 1 and the features of the data used are presented in Table 2. Furthermore, from the data that has been obtained proceed to the data preprocessing stage. At this stage, cleaning of empty values, duplicate values, and data normalization is carried out.

Table 1. Amount of Data on Each E-Commerce

No	Ecommerce Name	amount of data
1.	UK Retail	1000
2.	Olist Store	1070
3.	Pakistan Alnafi	1050

Table 2. Data Features in Each E-Commerce

No	Feature Name	Description
1.	customer_id	A unique value that describes a single user
2.	length	The value of the result of subtracting the date of the last transaction with the first
3.	recency	The result of subtracting the last transaction date by data analysis
4.	frequency	The number of transactions made by the user
5.	monetary	Total transactions made by the user

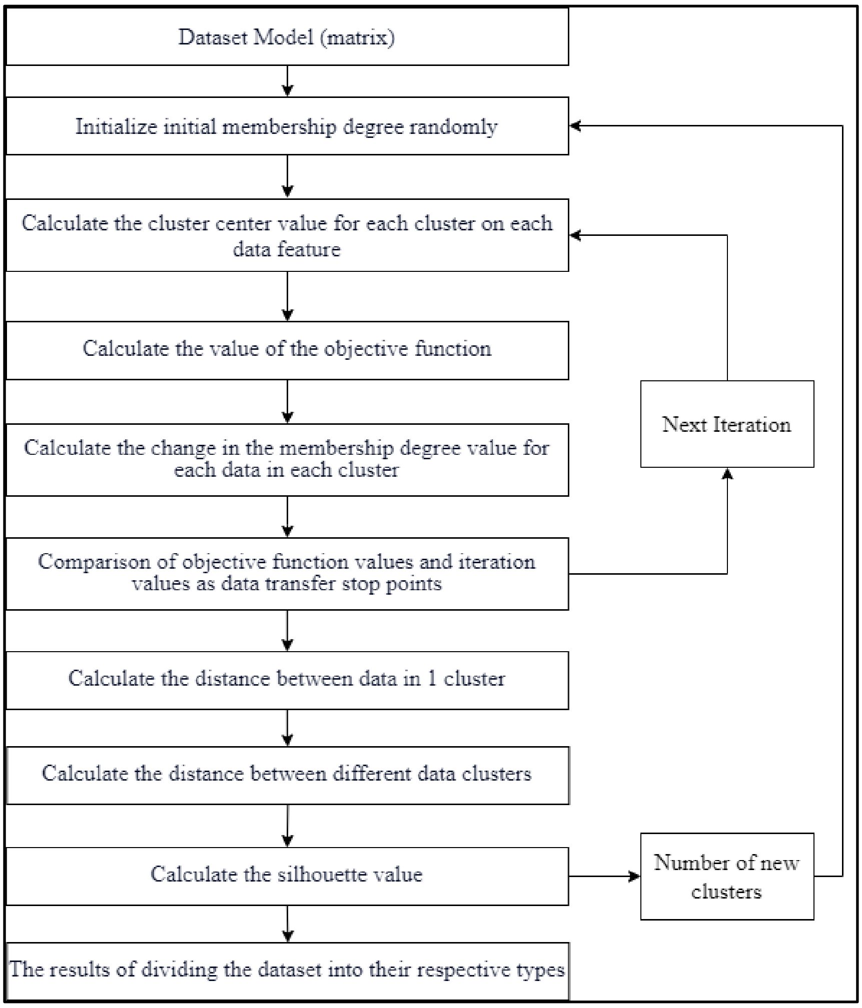

Figure 2. The Algorithm of Fuzzy-Silhouette Model

The fourth stage is implementing the Fuzzy-Silhouette model. The proposed model consists of a clustering model that predicts the number of types that fit the data and separates the data into the best type. The model was created by integrating the silhouette coefficient function in determining the best number of clusters into the Fuzzy C-Means algorithm. The combined model is called the Fuzzy-Silhouette model. Figure 2 shows the structure of the proposed model. The first stage in the proposed model is to initialize the initial membership degree randomly for each data. The degree of membership is a degree that indicates how likely a data can become a member of a cluster (Sanusi, Zaky, & Afni, 2019). In the proposed model, the initial number of clusters is 2. The initial random membership degree requirement for each data is equal to one if all membership degrees in each cluster are added up and cannot be the same between clusters. This is because each data may not be included in 2 clusters (Efiyah, 2014). Equation 1 shows the formula for the initial degree of membership.

$$\sum_{j=1}^{c} U_{ij} = 1 \quad (1)$$

Where, c = the number of cluster

Dataset models that already have membership degrees will calculate the centroid value. Centroid calculations are performed to see the distance between the data and the centroid in each cluster. The lesser the distance data to the centroid, the better data is clustered. In addition, the centroid is also one of the variables used in calculating the objective function of the model (Rahakbauw et al., 2017). The calculation formula of the centroid is illustrated in Equation 2.

According to Equation 2, the centroid is calculated for each feature in each cluster (V_{kj}). The value of w (Fuzziness level) that is used is 2. This value was chosen based on a study conducted by Gupta (2018) suggested the optimal level of fuzziness is at the value of 2, because in that value the best validity value of the cluster will be obtained. In the proposed model, the objective function is calculated using Equation 3. In that formula, the calculation of the objective function will be carried out in each iteration (Pt). The objective function is used to measure the accuracy of the membership degree value for each data and the value used to pass the threshold limit set in the study.

$$V_{kj} = \frac{\sum_{i=1}^{n} U_{ik}{}^{w} * X_{ij}}{\sum_{i=1}^{n} U_{ik}{}^{w}} \quad (2)$$

$$P_t = \sum_{i=1}^{n} \sum_{k=1}^{c} \left(\left[\sum_{j=1}^{m} \left(X_{ij} - V_{kj} \right)^2 \right] (U_{ik})^w \right) \quad (3)$$

Changes in the membership degree value are made to minimize the value of the objective function in the next iteration. Equation 4 shows the formula for calculating the updated value of the membership degree (Prabowo and Kurniawan, 2018). According to Equation 4, membership degree update for each data in each cluster (U_{ik}). The result of the updated membership degree value will determine the cluster for data. To determine whether the cluster is the data's best cluster, it will do validation that is carried out by subtracting two objective function values. If the reduction result is smaller than the threshold that is used, then the data is in the best cluster. However, if is not, it will repeat from the centroid process.

$$U_{ik} = \frac{\left[\sum_{j=1}^{m} (x_{ij} - V_{kj})^2\right]^{\frac{-1}{w-1}}}{\sum_{k=1}^{c} \left[\sum_{j=1}^{m} (x_{ij} - V_{kj})^2\right]^{\frac{-1}{w-1}}} \quad (4)$$

The next step in the Fuzzy-Silhouette model after all the data has clusters is to check the accuracy of the number of clusters used. The inspection is carried out by measuring the minimum intra-cluster distance to the maximum inter-cluster distance. If the value obtained is getting closer to one, then the number of clusters used in the data is the best (Anggraeni, 2015). Intra-cluster distance measurement is illustrated in equation 5. In this formula, distance calculations are performed for each data in its cluster using the Euclidean distance. Euclidean distance is a measure of the distance between 2 numerical data points (Aditya, Sari, & Padilah, 2020). Equation 6 shows the formula for calculating the distance between clusters for each data. Based on Equation 6, each data will calculate a different data distance from the cluster itself. From these results, the minimum distance obtained will be used to search for Silhouettes in checking the number of clusters (Edo, 2021). The formula for calculating the Silhouette value is illustrated in equation 7.

$$a(i) = \frac{1}{[A]-1} \sum j \in_{A, j \neq i} \; d(i,j) \quad (5)$$

$$b(i) = \frac{1}{[A]} \sum j \; C \; d(i,j) \quad (6)$$

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (7)$$

The next stage in the research is to evaluate the model. Model evaluation in this study was carried out using the Davied-Bouldin Index (DBI) level. The Davied-Bouldin Index (DBI) is an evaluation method that measures the level of validity and quality of clusters resulting from the clustering method. This DBI measurement examines the maximum distance level between clusters as well as tests the minimum intra-cluster distance level. If the distance between clusters is maximum, then the characteristics of each cluster will be seen so that the differences between clusters will also be seen (Wani & Riyaz, 2017). The smaller the DBI value obtained (non-negative >= 0), the better the cluster obtained from the clustering algorithm (Bates &Jugal, 2016). The DBI mathematical calculation formula is shown in Equation 8.

$$DBI = \frac{1}{k} \sum_{i=1}^{k} max_{i \neq j} \left(R_{i,j} \right) \quad (8)$$

The next stage in the research after getting the best results from the model used is to analyze the type of loyalty of e-commerce users from the clustering results. The analysis was performed using the standard deviation. The standard deviation describes the distribution of data within groups. For each cluster obtained, the standard deviation is calculated using the formula in Equation 9 for each numerical feature (Firdhausyah, Maulana, & Setiawan, 2021). Based on equation 9, the standard deviation (S) will be calculated using the feature data value (Xi) and the average. The results obtained according to equation 9 will be analyzed under the conditions, if the standard deviation for each feature in a cluster is higher than the average, an up arrow (↑) will be given for the abbreviation symbol for the feature. However, if it is lower, it will be given a downward arrow (↓) as a symbol. From these conditions, a feature symbol will be obtained to determine the type of loyalty (Hughes, 2012). There are 16 combinations of user loyalty types in e-commerce which are illustrated in table 3 (Monalisa, 2018; Nios, 2020).

$$S = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}} \quad (9)$$

Table 3. User Loyalty Type Combination

No	Feature Value	User Segment
1.	L↑R↓F↑M↑	*High Value Loyal Users*
2.	L↑R↓F↑M↓	*High Frequency Buying Users*
3.	L↑R↓F↓M↑	*Platinum Users*
4.	L↑R↑F↑M↑	*Potential Loyal Users*
5.	L↑R↑F↑M↓	*Potential High Frequency Users*
6.	L↑R↑F↓M↑	*Potential Consumption Users*
7.	L↓R↑F↑M↑	*High Value Lost Users*
8.	L↓R↑F↑M↓	*Frequency Lost Users*
9.	L↓R↑F↓M↑	*Consumption Lost Users*
10.	L↓R↑F↓M↓	*Uncertain Lost Users*
11.	L↓R↓F↑M↑	*High Value New Users*
12.	L↓R↓F↑M↓	*Frequency Promotions Users*
13.	L↓R↓F↓M↑	*Consumption Promotions Users*
14.	L↓R↓F↓M↓	*Uncertain New Users*
15.	L↑R↓F↓M↓	*High Consumption Cost Users*
16.	L↑R↑F↓M↓	*Low Consumption Cost Users*

RESULT AND DISCUSSION

Best Silhouette Coefficient Result

The implementation of the Fuzzy-Silhouette model is carried out after the dataset used goes through the data preprocessing stage. In the three datasets used, namely UK Retail, Olist Store, and Pakistan Alnafi, there is no change in the amount of data from the results of the preprocessing data, so the amount of data used is following table 1. The first stage in the implementation of the Fuzzy-Silhouette model is carried out by initializing the fixed variables used in this study. The fixed variables include the threshold value used is 0.0001, the number of iterations is 1000, and the level of fuzziness of the implemented model is 2. After initialization is carried out, the model performs the process of grouping types and the number of types that match according to the dataset used. The results released from the implementation of the model on the three datasets can be seen in table 4, where the silhouette values and the best number of clusters for each dataset are given in the table. The silhouette values presented in the table are the largest values generated from the Fuzzy-Silhouette model for each dataset entered. Based on the table, the distribution of data for each cluster and dataset can be seen in Figure 3. In this figure, you can see the distribution of data from the UK Retail dataset in section (a), the Olist Store dataset (b), and the Pakistan Alnafi dataset (c).

Table 4. Results of the Best Silhouette Value and Number of Clusters

No	Dataset name	Number of Clusters	Silhouette Value
1.	UK Retail	4	0.78
2.	Olist Store	2	0.595
3.	Pakistan Alnafi	3	0.858

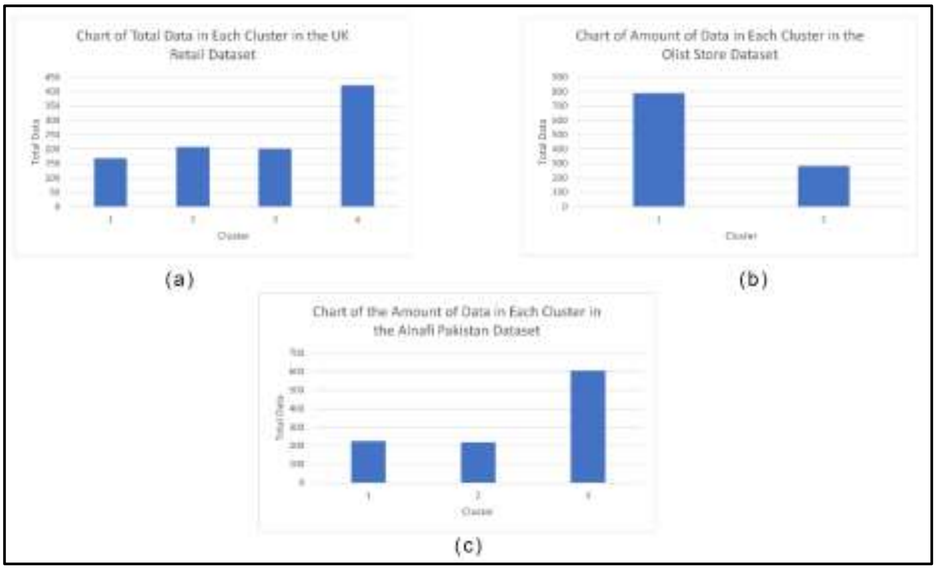

Figure 3. Distribution of the Amount of Data for Each Cluster in Each Dataset

Analyze The Type of Loyalty User

The type of user loyalty is analyzed using the standard deviation value of the grouping results which is the output of the Fuzzy-Silhouette model. In addition, the average standard deviation value for each dataset is also calculated. The results of calculating the standard deviation value of the UK Retail dataset can be presented in table 5, Olist Store in table 6, and Pakistan Alnafi in table 7.

Table 5. UK Retail Dataset Standard Deviation Calculation Results

Cluster	Length	Recency	Frequency	Monetary
1	0.0925	0.0360	0.1586	0.1379
2	0.1169	0.0882	0.0903	0.0890
3	0.0548	0.2289	0.0499	0.0748
4	0.0769	0.0333	0.0623	0.0630
Average	0.0853	0.0966	0.0901	0.0912

Table 6. Olist Store Dataset Standard Deviation Calculation Results

Cluster	Length	Recency	Frequency	Monetary
1	0.1064	0.0444	0.1499	0.1573
2	0.1533	0.2168	0.0350	0.0399
Average	0.1298	0.1306	0.0924	0.0986

Table 7. Alnafi Pakistan Dataset Standard Deviation Calculation Results

Cluster	Length	Recency	Frequency	Monetary
1	0.0561	0.2234	0.0037	0.0025
2	0.1073	0.0867	0.0113	0.0004
3	0.0701	0.0323	0.0424	0.0405
Average	0.0817	0.1550	0.0075	0.0014

Based on the three tables, an analysis of feature values is carried out with the provision that if the standard deviation result for each feature in a cluster is higher than the average, an up arrow (🡡) will be given for the abbreviation symbol of the feature. However, if it is lower, it will be given a downward arrow (🡣) as a symbol. From these conditions, a feature symbol will be obtained to determine the type of loyalty (Hughes, 2012). The results of the feature values obtained are compared with table 3, resulting in a type of user loyalty for each dataset used. Types of user loyalty from UK Retail can be seen in table 8, Olist Store in table 9, and Pakistan Alnafi is presented in table 10.

Table 8. Result Type UK Retail Dataset User Loyalty

Cluster	Feature Value	User Loyalty Type
1	L🡡R🡣F🡡M🡡	*High Value Loyal Users*
2	L🡡R🡣F🡡M🡣	*High Frequency Buying Users*
3	L🡣R🡡F🡣M🡣	*Uncertain Lost Users*
4	L🡣R🡣F🡣M🡣	*Uncertain New Users*

Table 9. Result Type Olist Store Dataset User Loyalty

Cluster	Feature Value	User Loyalty Type
1	L↓R↓F↑M↑	*High Value New Users*
2	L↑R↑F↓M↓	*Low Consumption Cost Users*

Table 10. Result Type Pakistan Alnafi Dataset User Loyalty

Cluster	Feature Value	User Loyalty Type
1	L↓R↑F↓M↑	*Consumption Lost Users*
2	L↑R↓F↑M↓	*High Frequency Buying Users*
3	L↓R↓F↑M↑	*High Value New Users*

Evaluation Model

The Davied-Bouldin Index (DBI) level in this study assesses the level of validity of the results of the automation of determining the loyalty type of e-commerce users based on the results of the Fuzzy-Silhouette model. The level of validity is measured based on the minimum distance between data in the intra-cluster with the maximum distance between the data in the inter-cluster. The results of calculating the DBI level for each dataset used are presented in table 11.

Table 11. DBI Level Evaluation Results

No	Dataset Name	DBI Rate
1.	UK Retail	0.31
2.	Olist Store	0.36
3.	Pakistan Alnafi	0.24

Based on the results of the research analysis, it can be seen that the Fuzzy-Silhouette model can automatically generate the number and classification of user loyalty types with good validity based on the DBI value. The smaller the DBI value obtained (non-negative >= 0), the better the cluster obtained from the clustering algorithm (Butsianto & Saepudin, 2020; Cahyo, Subekti, & Haris, 2022). In table 11, it can be seen that the DBI values obtained in the 3 datasets used are close to zero (0).

The results obtained in this study are in line with previous research which also revealed that the greater the Silhouette value that can be produced (closer to one (1) value), the more accurate the data grouping is carried out (Paembonan & Abduh, 2021; Hidayati, et al., 2021). The results of other studies also revealed that grouping data using Fuzzy C-Means was able to produce more accurate groupings compared to other algorithms (Wulandari & Yogantara, 2022; Widiyanto, 2019). So, based on some of the research results it can be said that Fuzzy-Silhouette can produce a good validity value in the automation of determining the type of user loyalty in e-commerce. In this study, there are limitations in determining the use of features from the dataset, so it is recommended to increase the number of features used to expand the combination of types of user loyalty in e-commerce.

CONCLUSSION

In this research, the Fuzzy C-Means method was proposed to combine with the silhouette function as an automated function for determining the best number of user loyalty types that are suitable for the dataset that is used. In the model, the number of clusters parameter that is used is set to start from 2 until a suitable number of clusters is found. The study performed a detailed experiment to analyze user loyalty type. Based on the results, the Fuzzy-Silhouette model in determining the type of loyalty of e-commerce users can be concluded in general that the results of the automation of the generated user loyalty types have good validity with consecutive validity values on the 3 research subjects used, namely 0.31, 0.36, and 0.24. The three values obtained are close to zero (0)) meaning that the inter-cluster distance has the maximum possible value and the distance between clusters has been kept to a minimum so that the characteristics of each cluster are visible. The results of this study can be used by business actors in optimizing the use of e-commerce as a marketing communication strategy. In this study, 4 numerical features were used in determining customer loyalty, in future research, it is hoped that it can increase the number of features used so that the range of type combinations obtained is increasingly diverse.

REFERENCES

[1] A. R. Irawan, "Strategi Komunikasi Pemasaran Melalui Media E-Commerce Dalam Meningkatkan Penjualan Produk Bangbrew," *Jurnal Ilmiah Mahasiswa Ilmu Sosial dan politik (JIMSIPOL),* pp. 1-15, 2022, [Online]. Available: http://jurnalmahasiswa.umsu.ac.id/index.php/jimsipol/article/view/1219/1259.

[2] G. Jayanti, "Penerapan Ecommerce Dalam Proses Komunikasi Pemasaran Di Toko East Kalimantan Center Samarinda," *eJournal Ilmu Komunikasi,* pp. 219-228, 2014, [Online]. Avai;able: https://ejournal.ilkom.fisip-unmul.ac.id/site/?p=1484.

[3] M. N. Rahiim, "E-Commerce: Studi Perlindungan Hukum Dalam Transaksi Jual Beli Online di Situs Bukalapak.com," Universitas Muhammadiyah, Surakarta, 2018.

[4] A. S. Nasution, "Analisis Teori Kebangkrutan Dengan Metode ALTMAN Z-SCORE Pada Nokia Corporation Periode 2015-2017," Universitas Pembangunan Panca Budi, Medan, 2019.

[5] B. Borhanuddin, "Nokia: An Historical Case Study," *Electronic Journal of Computer Science and Information Technology (eJCSIT),* pp. 1-14, 2016, [Online]. Available: http://ejcsit.uniten.edu.my/index.php/ejcsit/article/view/95.

[6] B. H. Alireza and S. David, Data segmentation and Model Selection for Computer Vision: a statistical approach, New York: Springer Science and Business Media, 2012.

[7] R. Siagian, P. Sirait and A. Halima, "E-Commerce Customer Segmentation Using K-Means Algorithm and Length, Recency, Frequency, Monetary Model," *Journal of Informatics and Telecommunication Engineering,* pp. 1-10, 2021, doi: 10.31289/jite.v5i1.5182.

[8] N. Agustina and Prihandoko, "Perbandingan Algoritma K-Means Dengan Algoritma Fuzzy C-Means Untuk Clustering Tingkat Kedisiplinan Kinerja Karyawan," *Jurnal Rekayasa Sistem dan Teknologi Informasi,* pp. 621-626, 2018, doi: 10.30998/faktorexacta.v8i3.327.

[9] R. Siringoringo and Jamaludin, "Improving the Performance of Fuzzy C-Means Clusters in Sentiment Clustering Using Particle Swarm Optimization," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK),* pp. 349-354, 2019, doi: 10.25126/jtiik.2018561090.

[10] M. Mohammdian-khoshnoud, A. R. Soltanian, A. Dheghan, M. Farhadian, Sol and Deg, "Optimization of fuzzy c-means (FCM) clustering in cytology image segmentation using the gray wolf algorithm," *BMC Molecular and Cell Biology,* pp. 2-9, 2022, doi: 10.1186/s12860-022-00408-7.

[11] J. Perez-ortega, S. S. Roblero-agruilar, N. N. Almanza-Ortega, J. F. Solis, C. Zavala-Diaz, Y. Hernandez and V. Landero-Najera, Hybrid Fuzzy C-Means Clustering Algorithm Oriented to Big Data Realms, Multidisciplinary Digital Publishing Institute, 2022, doi: 10.3390/axioms11080377.

[12] W. Sanusi, A. Zaky and B. N. Afni, "Analisis Fuzzy C-Means dan Penerapannya Dalam Pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan Berdasarkan Faktor-faktor Penyebab Gizi Buruk," *Journal of Mathematics, Computations, and Statistics,* pp. 47-54, 2019, doi: 10.35580/jmathcos.v2i1.12458.

[13] D. Rahakbauw, Ilwaru and Hahury, "Implementasi Fuzzy C-Means Clustering Dalam Penentuan Beasiswa," *Jurmal Ilmu Matematika dan Terapan,* pp. 1-11, 2017, doi: 10.30598/barekengvol11iss1pp1-12.

[14] E. Prabowo and R. Kurniawan, "Optimasi Algoritma Fuzzy Clustering dengan Mengggunakan Algoritma Forest Optimization," *Information System Development (ISD),* vol. 4, pp. 1-6, 2019, [Online]. Available: https://ejournal-medan.uph.edu/index.php/isd/article/view/214/83.

[15] W. Anggraeni, "Penentuan Nilai Pangkat Pada Algoritma Fuzzy C-Means," *Factor Exacta,* vol. 8(3), no. 1979-276x, pp. 266-278, 2015, doi: 10.30998/faktorexacta.v8i3.327.

[16] A. Aditya, B. N. Sari and T. N. Padilah, "Comparison analysis of Euclidean and Gower distance measures on k-medoids cluster," *Jurnal Teknologi dan Sistem Komputer,* pp. 1-7, 2020, doi: 10.14710/jtsiskom.2021.13747.

[17] D. F. Edo, "Perbandingan Metode Elbow dan Silhouette Coefficient pada Algoritma K-Medoids Clustering Dalam Pengelompokan Penilaian Dosen Dalam Pembelajaran Berdasarkan Evaluasi Mahasiswa (Studi Kasus : Jurusan JTEIF Institut Teknologi Sumatera)," Institut Teknologi Sumatera, Lampung Selatan, 2021.

[18] M. A. Wani and R. Riyaz, "A novel point density based validity index for clustering gene expression datasets," *International Journal of Data Mining and Bioinformatics,* vol. 17(1), p. 66–84, 2017, doi: 10.1504/IJDMB.2017.084027.

[19] A. Bates and K. Jugal, "Counting clusters in twitter posts," in *Proceedings of the 2nd International Conference on Information Technology for Competitive Strategies*, Udaipur, 2016, [Online]. Available: http://www.cs.uccs.edu/~jkalita/papers/2016/BatesAndrewSoDA2016.pdf.

[20] H. Firdhausyah, R. Maulana and E. Setiawan, "Sistem Pendeteksi Atrial Fibrilasi Berdasarkan Fitur Mean, Median, Standar Deviasi, Min, dan Maks Interval RR menggunakan Metode K-NN," *Jurnal Pengembangan Teknologi Informasi dan Komputer,* pp. 2327-2333, 2021.

[21] A. M. Hughes, Strategic Database Marketing, Chicago: Probus Publishing Company, 2012.

[22] S. Monalisa, "Klasterisasi Customer Lifetime value dengan Model LRFM Menggunakan Algoritma K-Means," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK),* pp. 247-252, 2018, doi: 10.25126/jtiik.201852690.

[23] Nios, Business Studies (Chapter 16: Sales Promotion and Personal Selling), India: The National Institute Of Open Schooling, 2020.

[24] S. Butsianto and N. Saepudin, "Penerapan Data Mining Terhadap Minat Siswa Dalam Mata Pelajaran Matematika Dengan Metode K-Means," *Jurnal Nasional Komputasi dan Teknologi Informasi,* pp. 51-59, 2020, [Online]. Available: https://ojs.serambimekkah.ac.id/jnkti/article/view/2008/pdf.

[25] K. N. Cahyo, A. Subekti and M. Haris, "Analisis Performa dan Kecepatan Komputasi Algoritma K-Means dan K-Medoids Pada Text Clustering," *Jurnal ilmiah dan Elektronika Komputer,* pp. 444-464, 2022, doi: 10.51903/pixel.v15i2.931.

[26] S. Paembonan and H. Abduh, "Penerapan Metode Silhouette Coefficient Untuk Evaluasi Clustering Obat," *Jurnal Ilmiah Ilmu-Ilmu Teknik,* pp. 48-54, 2021, [Online]. Available: https://ojs.unanda.ac.id/index.php/jiit/index.

[27] R. Hidayati, A. Zubair, A. H. Pratama and L. Indayana, "Analisis Silhouette Coefficient Pada 6 Perhitungan Jarak K-Means Clustering," *Jurnal Teknologi Informasi,* pp. 186-197, 2021, doi: 10.33633/tc.v20i2.4556.

[28] L. Wulandari and B. O. Yogantara, "Algorithms Analysis of K-Means and Fuzzy C-Means for Clustering Countries Based On Economy and Health," *Faktor Exacta,* pp. 109-116, 2022, doi: 10.30998/faktorexacta.v15i2.12106.

