Jurnal CoreIT, Vol.11, No.2, December 2025
ISSN 2460-738X (Print)
ISSN 2599-3321 (Online)

Clustering of Halal MSME Aid Recipients: Uncovering Patterns and Characteristics Using the K-Medoids Method

Yelfi Vitriani[1], Siska Kurnia Gusti[2*]
[1]Dept. of Informatics Engineering, Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia
yelfi.vitriani@uin-suska.ac.id, siskakurniagusti@uin-suska.ac.id*

Abstract. The rapid growth of the halal industry has strengthened the strategic role of Micro, Small, and Medium Enterprises (MSMEs) in meeting market expansion. However, the absence of structured insights regarding the characteristics and patterns of halal MSME aid recipients has hindered the formulation of effective and targeted support programs. This study aims to identify the clustering patterns of halal MSME beneficiaries in Indonesia using the K-Medoids algorithm optimized with Principal Component Analysis (PCA). A total of 129 MSME datasets were collected through validated questionnaires consisting of demographic variables, aid history, business performance, and operational challenges. Preprocessing included data cleaning, transformation, and dimensionality reduction using PCA. The optimal PCA dimension was determined as two components based on the Davies-Bouldin Index (0.1737). K-Medoids clustering produced three optimal clusters validated using Silhouette (0.4602), Davies-Bouldin Index (0.7861), and Elbow Method (K>=3). Each cluster shows distinctive characteristics in income range, business legality, type of aid received, challenges, and performance outcomes. The novelty of this research lies in the application of PCA-optimized K-Medoids for halal MSME segmentation, providing insightful foundations for evidence-based policymaking.

Keywords : Halal, K-Medoids, PCA, MSMEs.

Received November 2025 / **Revised** November 2025 / **Accepted** December 2025

INTRODUCTION

Over the past decade, the halal industry has increasingly become a focal point of the global economy. The growing demand for halal products, driven by heightened awareness of religious values, hygiene, and safety, has significantly reshaped the business landscape. Moreover, the Micro, Small, and Medium Enterprises (MSMEs) sector within this industry has emerged as a critical component in meeting the rapidly expanding market demand. MSMEs play a vital role in Indonesia's economy, contributing more than 60% to the national Gross Domestic Product (GDP), equivalent to approximately IDR 8,573 trillion annually. Furthermore, MSMEs provide employment for 97% of the Indonesian workforce, or around 116 million people [1]. In response to this growth, both government and non-governmental organizations have implemented various assistance programs aimed at supporting MSMEs.

In Indonesia, halal product oversight is primarily managed by the Halal Product Assurance Organizing Agency (BPJPH), established under Law No. 33 of 2014 concerning Halal Product Assurance [2]. BPJPH holds the principal responsibility for halal certification and supervision across the country. In accordance with the law, BPJPH collaborates with the Halal Inspection Agency (LPH) and the Indonesian Ulema Council (MUI) in administering the Halal Product Assurance (JPH) system [3]. Additionally, the law provides MSME actors with access to financial assistance sourced from the State Budget (APBN) and Regional Government Budget (APBD) to facilitate the halal certification process.

Despite these initiatives, a substantial gap remains in understanding the patterns and characteristics of halal MSME aid recipients. The lack of detailed information regarding variations and similarities among beneficiary groups poses challenges to developing more targeted and effective support strategies. Addressing this gap, the present study focuses on clustering halal MSME aid recipients using the K-Medoids method.

The K-Medoids method, also referred to as PAM (Partitioning Around Medoids), is a data clustering technique that partitions a set of objects into k clusters. The algorithm selects specific data points as medoids, which represent the centers of each cluster [4]. Unlike K-Means, which uses the mean as the cluster center, K-Medoids chooses actual data points, resulting in a more accurate representation of cluster centers [5]. This method has been

DOI: 10.24014/coreit.v11i2.38634

applied in various studies, such as clustering provinces based on educational indicators and educational facilities [6]. K-Medoids is known for its efficiency and its ability to overcome the limitations of K-Means in determining optimal cluster centers [7]. Research by Syed Ali Abbas on birth data clustering in the city of Muzzaffarabad demonstrates that K-Medoids outperforms K-Means [8]. William Ramdhan (2022) conducted a study titled *"Clustering Algorithm-Based Pandemic Multicluster Framework Analysis: K-Means and K-Medoids"*, which investigates the effectiveness of a multicluster analytical framework in identifying key variables influencing pandemic conditions. The study utilizes the Davies–Bouldin Index (DBI) to evaluate clustering performance generated by the K-Means and K-Medoids algorithms [9]. Nurhayati conducted a study entitled *"Analysis of K-Means and K-Medoids Performance Using Big Data Technology"*, which compares the performance of the K-Means and K-Medoids clustering methods. The results indicate that K-Medoids outperforms K-Means in terms of accuracy, achieving an average accuracy of 63.24%, compared to 52.11% for K-Means. Additionally, K-Medoids demonstrates superior execution time performance, with an average processing time of 3.1 ms, whereas K-Means requires 3.45 ms [10]. Another study by Mahdi Hashemzadeh shows that the K-Medoids method is reliable in identifying potential fire areas through video recordings [11].

A more comprehensive understanding of these patterns and characteristics is expected to generate more targeted recommendations and adaptive solutions for supporting halal MSMEs. Furthermore, this study contributes significantly to the formulation of policies, support strategies, and guidelines for government entities, related institutions, and other stakeholders, with the aim of strengthening the halal MSME sector, promoting sustainable economic growth, and meeting the increasingly dynamic demands of the market.

K-MEDOIDS

The K-Medoids method is a clustering analysis technique that aims to group data into clusters based on their similarities [12, 13, 14]. Unlike K-Means, which uses the arithmetic mean as the cluster center, K-Medoids employs actual data points as medoids, thereby avoiding the instability that may arise in K-Means due to its sensitivity to the initial selection of centroids. The fundamental difference between K-Medoids and K-Means lies in the choice of cluster centers: K-Medoids selects representative data points (medoids), whereas K-Means uses mean values as cluster centers [15, 16].

The primary steps of the K-Medoids algorithm are as follows [6]:

1) Initialization: Select K random data points as the initial medoids.
2) Cluster Formation: Assign each data point to the nearest medoid.
3) Optimization: For each medoid, evaluate the total distance of all data points associated with it. Replace the medoid with another data point if doing so results in a reduction of the total distance.
4) Iteration: Repeat steps 2 and 3 until convergence (i.e., no medoid changes or the maximum number of iterations is reached)

The general formula for calculating the distance between two points in a data space (e.g., using Euclidean distance for numerical data) is as follows:

For two points P = (p1,p2,...,pn), P = (p_1, p_2, ..., p_n), PP =(p1,p2,...,pn) and Q =(q1,q2,...,qn) in n-dimensional space:

$$d_{ij} = \sqrt{\sum_{a=1}^{p}}(x_{ia} - x_{ja})^2 \qquad (1)$$

Explanation:

d_{ij} = distance between objects i and j

x_{ia} = value of object i for variable a

x_{ja} = value of object j for variable a

p = total number of observed variables

In the context of the K-Medoids method, an essential step is to minimize the total distance between each data point and the medoid to which it is assigned. Thus, the formula for calculating the total distance from a data point *i* to a medoid *j* is expressed as follows:

$$T(i,j) =. \sum_{k=1}^{n}(Xi, Mj) \qquad (2)$$

Where:

Xi represents the *i-th* data point.

Mj represents is the *j-th* medoid.

$D(Xi,Mj)$ represents the distance between data point X_1 and medoid X_j, which can be computed using various distance metrics (such as Euclidean, Manhattan, and others). The K-Medoids algorithm focuses on selecting the most appropriate medoids to optimize cluster formation. Through iterative updates and evaluation of total distances, the algorithm seeks to identify the medoids that minimize the overall distance between the data points and their assigned medoids.

To identify the appropriate number of clusters in the K-Medoids method, this study employs several evaluation techniques, including Silhouette Coefficient, Davies–Bouldin Index (DBI) and Elbow Method.

SILHOUETTE COEFFICIENT

The Silhouette Coefficient is a technique used to evaluate the quality and appropriateness of object placement within a cluster [17]. It combines two key concepts: *cohesion* and *separation*. Cohesion measures how closely related an object is to other objects within the same cluster, while separation measures the distance or difference between that cluster and other clusters. The steps for calculating the Silhouette Coefficient are as follows:

1. Compute the average distance between an object—let's say object *i*—and all other objects within the same cluster.
$$a(i) = \sum j \in_{A,j \neq i} d(i,j) \qquad (4)$$
2. Calculate the average distance between object *i* and all objects in the other clusters, then take the smallest value.
$$d(i,C) = \sum j \in C \; d(i,j) \qquad (5)$$
3. The Silhouette Coefficient value is:
$$\mathrm{s(i)} = \frac{b(i)-a(i)}{\max\,(a(i),b(i))} \qquad (6)$$

The results obtained using the Silhouette Coefficient method range between –1 and 1. The average Silhouette Coefficient value of all objects within a cluster represents how well the data points are grouped within that cluster. The closer the average Silhouette Coefficient value is to 1, the better the quality of the clustering, indicating that the objects are highly similar within the same cluster. Conversely, if the value approaches –1, the clustering quality becomes poorer, suggesting that the objects may have been placed in the wrong cluster. The following are the criteria for interpreting Silhouette Coefficient values [12]:

Silhouette Coefficient Value Scale

Table 1. Silhouette coefficient value scale

SILHOUETTE COEFFICIENT	INTERPRETASI
$0.7 < SC \leq 1.0$	Stronge Structure
$0.5 < SC \leq 0.7$	Mediun Structure
$0.25 < SC \leq 0.5$	Weak Structure
$SC \leq 0.25$	No Structure

METHOD

This study uses data collected directly from questionnaires distributed to members of the MSME community in Riau Province, Indonesia. The questionnaire consists of several questions that have been analyzed and validated by experts. The contents of the questionnaire include: MSME Profile Data, Business Assistance, and Halal MSME Performance Measurement (covering Revenue, Product and Service Quality, Market Access, and Networking and Collaboration).

Table 2. Halal MSME performance measurement scale

PATTERNS AND CHARACTERISTICS OF MSMEs RECEIVING ASSISTANCE						
No	**Halal MSME Performance Measurement**	**Level of Presence/Agreement**				
		1	**2**	**3**	**4**	**5**
A. Sales Revenue						

PATTERNS AND CHARACTERISTICS OF MSMEs RECEIVING ASSISTANCE						
No	**Halal MSME Performance Measurement**	**Level of Presence/Agreement**				
		1	**2**	**3**	**4**	**5**
1	The number of products sold increases every month	1	2	3	4	5
2	The number of customers increases every month	1	2	3	4	5
3	Customer orders are increasing	1	2	3	4	5
B. Product and Service Quality						
4	After receiving assistance, my product quality improved	1	2	3	4	5
5	After receiving assistance, I have new products to sell	1	2	3	4	5
6	The number of products I produce increases every month	1	2	3	4	5
7	My production process is becoming more efficient	1	2	3	4	5
8	The products I produce are environmentally friendly	1	2	3	4	5
9	I have updated my product packaging to increase customer appeal	1	2	3	4	5
10	Customer service has become better and friendlier	1	2	3	4	5
C. Market Access						
11	My market share is expanding	1	2	3	4	5
12	I also sell products online through marketplaces and social media	1	2	3	4	5
13	Online sales are showing an increase in the number of customers	1	2	3	4	5
D. Networking and Collaboration						
14	I can easily collaborate with business partners	1	2	3	4	5
15	I am a member of an MSME community	1	2	3	4	5
16	Collaboration with business partners helps increase product sales	1	2	3	4	5

From the measurement scale data of Halal MSME Performance Assessment, a questionnaire was developed that produced 29 parameters. The data obtained from the questionnaire will then be processed using the K-Medoids clustering algorithm. The 29 parameters are as follows:

Table 3. Parameters used

No	Variable	Code
1	Gender	A
2	Age	B
3	Highest Education Level	C
4	Employment Status	D
5	Business Age	E
6	Type of Business	F
7	How many employees do you currently have?	V1
8	Monthly revenue (turnover)?	V2
9	Certifications/business permits owned	V3
10	If yes, specify the type of assistance received	V4
11	Source of assistance	V5
12	Biggest challenges in running a halal business	V6
13	Types of assistance needed	V7
14	Sales increase every month	V8
15	Customer count increases every month	V9
16	Customer orders are increasing	V10
17	Assistance improved product quality	V11
18	Assistance helps create new products	V12

No	Variable	Code
19	Production volume increases every month	V13
20	Production process becomes more efficient	V14
21	Products are environmentally friendly	V15
22	Packaging updated to increase customer appeal	V16
23	Customer service becomes better and friendlier	V17
24	Market share expands	V18
25	Selling online via marketplace and social media	V19
26	Online sales increase in customer count	V20
27	Easy collaboration with business partners	V21
28	Membership in MSME community	V22
29	Collaboration increases product sales	V23

IMPLEMENTATION OF K-MEDOIDS CLUSTERING

A total of 129 questionnaire responses were collected. The next stage is preprocessing, which involves examining all entries to identify any missing or inconsistent data. Several techniques can be applied to handle such issues; in this study, the mode value of each variable was used to fill missing or inconsistent entries. Afterward, the data underwent a transformation process based on the previously conducted analysis. After completing the data preprocessing stage, the next step involves optimizing the variables to be used in the analysis. This study employs Principal Component Analysis (PCA) as the optimization method. Prior to applying PCA, an additional optimization procedure is conducted to identify the most suitable number of dimensions (variables) that should be retained for the subsequent clustering process.

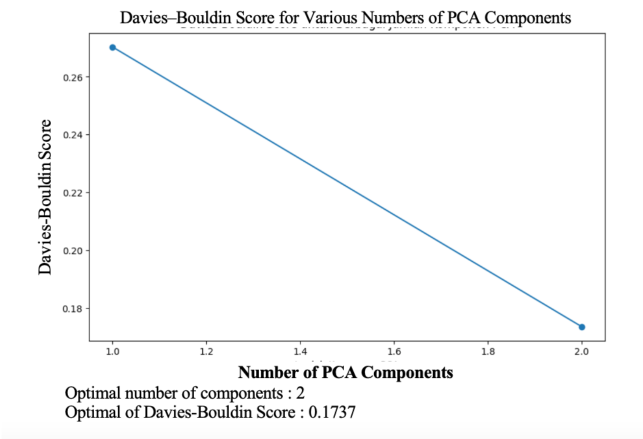


Figure 1. Determining the optimal number of dimensions using PCA

Based on the calculations, the analysis indicates that the optimal dimensionality is two components, yielding a Davies–Bouldin Index (DBI) score of 0.1737. Once the optimal number of dimensions is established, the K-Medoids algorithm is applied to perform the clustering process. The transformed feature set produced by the PCA serves as the input for the K-Medoids clustering procedure.

RESULT AND DISCUSSION

Using the K-Medoids method on the 129 input data points, and based on the PCA-reduced features consisting of 2 components with a DBI score of 0.01737, the analysis produced three clusters, summarized as follows:

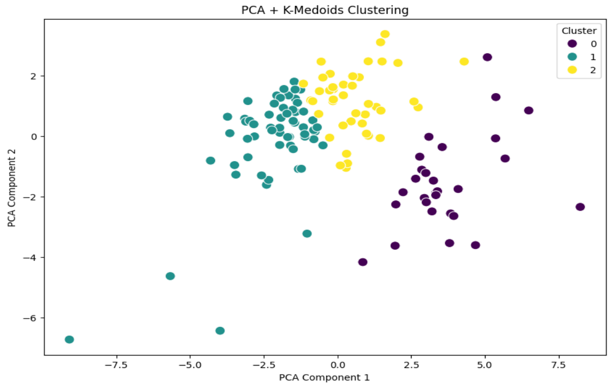


Figure 2. K-Medoids clustering results

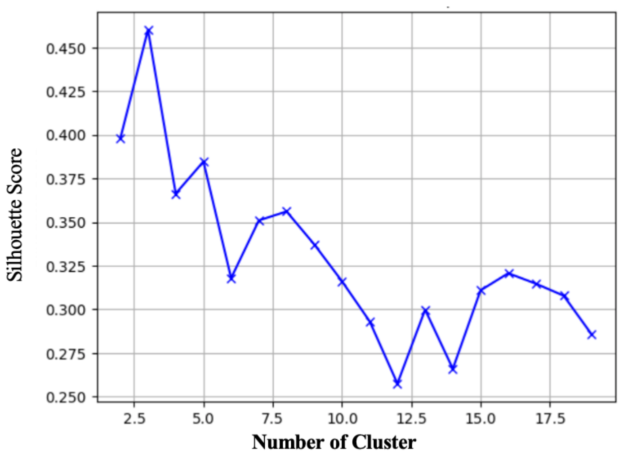
The clustering process resulted in three clusters, with the distribution of data points as follows:

Table 4. Number of data points in each cluster

Cluster	Data
0	44
1	30
2	55
Total Number	**129**

A. Silhouette Evaluation

Based on the evaluation using the Silhouette method, the optimal number of clusters is k=3, with a Silhouette score of 0.4602. Figure 4.5 illustrates the implementation of the Silhouette analysis:



Silhouette Score: 0.4602
For k:3

Figure 3. K Optimal for silhouette evaluation

B. Davies–Bouldin Index (DBI) Evaluation

In the evaluation using the DBI method, the optimal number of clusters was found to be 3**,** with a DBI value of 0.7861. In DBI analysis, a lower DBI value indicates better clustering quality. Based on the graph shown in Figure 4, it can be concluded that the clustering solution with 3 clusters provides the best performance.

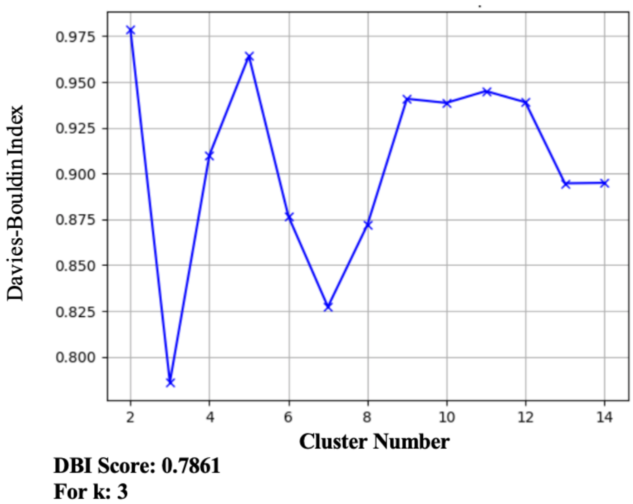


Figure 4. K Optimal for DBI evaluation

C. Elbow Evaluation

In this evaluation, the optimal value of K was found to be 3. Figure 5. below illustrates the application of the elbow method:

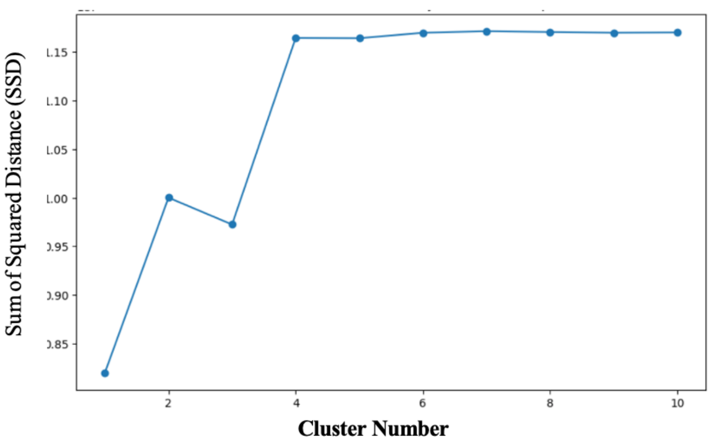


Figure 5. Elbow evaluation

Based on Figure 5, the optimal number of clusters is K = 3 (with a possible alternative at K = 4). This indicates that the formation of three clusters is sufficient to represent the underlying data structure, while adding more clusters does not provide a significant improvement in clustering quality.

CONCLUSION

The application of the K-Medoids method produced three distinct clusters, each exhibiting different characteristics. The clustering results were validated by determining the optimal number of clusters, which was found to be K = 3. The quality of the clustering was subsequently evaluated using three validation metrics: the Silhouette Score (0.4602), the DBI (0.7861), and the Elbow Method (K>=3).

Analysis of the three clusters indicates that Halal MSME actors demonstrate diverse patterns and characteristics. Cluster 1 is predominantly composed of *adult male sellers* with a monthly revenue below 5 million IDR. Their primary challenge relates to business management practices. Cluster 2 is largely dominated by *adult female entrepreneurs* who typically also work within their own businesses. Their major obstacles include market competition and the need for financial capital to operate and expand their enterprises. Cluster 3 consists mainly of *young adult males* with a higher revenue range of 5–10 million IDR per month. Their main requirement is improved access to market opportunities to enhance sales and customer reach.

All three clusters share a common need for business capital assistance. However, capital support does not uniformly translate into increased sales or customer growth across clusters. Only Cluster 2 shows noticeable improvement, whereas Clusters 1 and 3 experience relatively minimal increases in product sales and customer numbers. All clusters indicate some growth in online sales. Despite this, many MSME actors across the three clusters are not yet involved in MSME communities and continue to face difficulties in establishing collaborations with business partners, resulting in limited or restricted market reach.

REFERENCES

[1] A. Anastasya, "Data UMKM, Jumlah dan Pertumbuhan Usaha Mikro, Kecil, dan Menengah di Indonesia," 8 6 2023. [Online]. Available: https://ukmindonesia.id/baca-deskripsi-posts/data-umkm-jumlah-dan-pertumbuhan-usaha-mikro-kecil-dan-menengah-di-indonesia. [Accessed 14 12 2023].

[2] B. P. J. P. H. (BPJPH), "Tentang Badan Penyelenggara Jaminan Produk Halal (BPJPH)," Kemenag RI, 2023. [Online]. Available: http://bpjph.halal.go.id/detail/tentang-bpjph. [Accessed 18 November 2023].

[3] M. Khoeron, "Bagaimana Sinergi BPJPH, LPH, dan MUI dalam Sertifikasi Halal? Ini Penjelasan Kemenag," 15 3 2022. [Online]. Available: https://www.kemenag.go.id/pers-rilis/bagaimana-sinergi-bpjph-lph-dan-mui-dalam-sertifikasi-halal-ini-penjelasan-kemenag-t5j3dq. [Accessed 15 12 2023].

[4] M. Azmi, A. A. Putra, D. Vionanda and A. Salma, "Comparison the Performance of K-Means and K-Medoids Algorithms in Grouping Regencies/Cities in Sumatera Based on Poverty Indicators," *UNP Journal of Statistics and Data Science,* vol. 1, no. 2, pp. 59-66, 2023.

[5] F. R. Iskak, "Implementasi Algoritma K-Medoids dan K-Errorr untuk Pengelompokan Kabupaten/Kotadi Provinsi Jawa Tengah Berdasrkan Jumlah Produksi Peternakan tahun 2020," *Jurnal Gaussian,* vol. 1, no. 3, 2022.

[6] A. P. Fialine, D. A. Alodia, D. Endriani and E. Widodo, "Implementasi Metode K-Medoids Clustering untuk Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan," *SEPREN: Journal of Mathematics Education and Applied,* vol. 2, no. 2, 2021.

[7] U. Izzaty and Rahmi, "Pengklasteran di kabupaten/Kota di Provinsi Sumatera Barat Berdasarkan Indikator Kesejahteraan Masyarakat dengan Validitas Koefisien Silhouette," *Jurnal Matematika UNAND,* vol. 9, no. 2, pp. 192-198, 2020.

[8] S. A. Abbas, A. Aslam, . A. U. Rehman, S. Arif and . S. Z. H. Kazm, "K-Means and K-Medoids: Cluster Analysis on Birth Data Collected in City Muzaffarabad, Kashmir," *IEEE Access,* vol. 8, pp. 151847 - 151855, 2020.

[9] W. Ramdhan, O. S. Sitompul, E. B. Nababan and Sawaluddin, "Clustering Algorithm-Based Pandemic Multicluster Framework Analysis: K-Means and K-Medoids," in *2022 IEEE International Conference of Computer Science and Information Technology (ICOSNIKOM)*, Laguboti, North Sumatra, Indonesia, 2022.

[10] Nurhayati, N. S. Sinatrya, L. K. Wardhani and Busman, "Analysis of K-Means and K-Medoids's Performance Using Big Data Technology," in *2018 6th International Conference on Cyber and IT Service Management (CITSM)*, Parapat - Indonesia, 2018.

[11] A. Z. Mahdi Hashemzadeh, "Fire detection for video surveillance applications using ICA K-medoids-based color model and efficient spatio-temporal visual features," *Expert systems with applications,* 2019.

[12] B. Wira, A. E. Budianto and A. S. Wiguna, "Impelementasi Metode K-Medoids Clustering untuk Mengetahui Pola Pemilihan Program Mahasiswa Tahun 2018 di Universitas Kanjuruhang Malang," *Jurnal Terapan Sains & Teknologi Vol. 1, No. 3, 2019 Universitas Kanjuruhan Malang,* vol. 1, no. 3, 2019.

[13] S. U. Afifah, A. Fariza and A. Basofi, "Spatial-Temporal Visualization of Tuberculosis Vulnerability in East Java, Indonesia, Using the K-medoids Clustering Algorithm," in *International Electronics Symposium (IES)*, Surabaya, Jawa Timur, 2022.

[14] L. Tang, Z. Li and U. d. A. F. Jin-Kao Hao LERIA, "Solving the Single-Row Facility Layout Problem by K-Medoids Memetic Permutation Group," *IEEE Transactions on Evolutionary Computation ,* vol. 27, no. 2, pp. 251 - 265, 2023.

[15] T. Wang, Q. Li, D. J. Bucci, Y. Liang, B. Chen and P. K. Varshney, "K-Medoids Clustering of Data Sequences With Composite Distributions," *IEEE Transactions on Signal Processing ,* vol. 67, no. 8, pp. 2093 - 2106, 2019.

[16] J. A. d. M. Brito, G. S. Semaan and A. C. Fade, "An Effective VNS Algorithm for k-Medoids Clustering Problem," *IEEE Latin America Transactions,* vol. 20, no. 5, pp. 710 - 717, 2022.