Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.3, June 2026, pp. 2476~2488
e-ISSN: 2548-6861

Clustering of Food Security Levels in North Aceh Using K-Medoids

Tarisha Zhafira [1], Bustami [2], Nunsina [3]*
* Informatics Engineering, Malikussaleh University
tarisha.220170124@mhs.unimal.ac.id [1], bustami@unimal.ac.id [2], nunsina@unimal.ac.id [3]

Article Info

Article history:

Received 2026-04-28
Revised 2026-05-13
Accepted 2026-05-25

Keyword:

Clustering,
Davies-Bouldin Index,
Food Security,
K-Medoids,
Silhouette Coefficient.

ABSTRACT

Food security is essential for ensuring food availability, accessibility, and utilization. This study applies a multidimensional indicator framework covering social, economic, and infrastructure aspects to cluster regions in North Aceh Regency, addressing limitations of previous studies that primarily focus on agricultural production indicators and lack policy-oriented analysis. The analysis uses 2023 data from 27 sub-districts at the village level, comprising 852 villages. The indicators include: (1) the ratio of agricultural land area to total population, (2) the ratio of food supply facilities and infrastructure to households, (3) the ratio of population with the lowest welfare status to total population, (4) the proportion of villages without adequate transportation access via land, water, or air, (5) the ratio of households without access to clean water, and (6) the ratio of population per health worker relative to population density. Data processing involves preprocessing, normalization, and K-Medoids clustering, evaluated using the Davies–Bouldin Index (DBI) and Silhouette Coefficient (SC). The results identify six clusters: highly food insecure (C1) with 63 villages, food insecure (C2) with 92 villages, moderately food insecure (C3) with 179 villages, moderately food secure (C4) with 42 villages, food secure (C5) with 49 villages, and highly food secure (C6) with 427 villages. Most villages fall within moderately food insecure to highly food secure categories, indicating disparities in food security distribution. The DBI value of 3.085 indicates moderate cluster compactness, while the SC value of 17.75% suggests weak separation between clusters. These findings provide policy recommendations for targeted and equitable food security interventions.

This is an open access article under the CC–BY-SA license.

I. Introduction

Food security is a crucial aspect in ensuring the availability, accessibility, and proper utilization of food for the population. Meeting food needs plays a fundamental role in improving the quality of human resources and supporting socio-economic development [1]. At the global level, food security has become a major concern within development agendas, particularly since the emergence of food and economic crises that have exposed the vulnerability of global food systems to various shocks. In this context, food security is considered a key foundation for social stability, public health, and sustainable economic growth. According to the Food and Agriculture Organization (FAO), food security is not only related to the availability of food but also involves the ability of individuals and nations to access, utilize, and maintain a stable food supply over time [2]. In recent years, the condition of food security in Indonesia has faced significant challenges, as indicated by the decline in its ranking in the Global Food Security Index [3]. This shows that food security is not only related to production availability but also includes aspects of accessibility and sustainable food utilization. Therefore, both central and regional governments have the responsibility to ensure the availability, accessibility, and utilization of nutritious food in order to prevent food vulnerability [4]. This role aligns with the concept of food security, which emphasizes the importance of availability, access, utilization, and stability of food in ensuring sustainable food systems, as well as ensuring that every individual has adequate access to safe and nutritious food [5].

North Aceh Regency is one of the regions in Aceh Province with significant agricultural potential. However, it still faces challenges in the equitable distribution and sustainability of food security. Disparities between regions are clearly visible, where some areas have fertile land and high agricultural production, while others face limitations in land, access to clean water, and inadequate agricultural infrastructure. These conditions lead to uneven food distribution and even contribute to higher food prices in certain areas. In addition, natural factors such as floods and droughts affect agricultural productivity and food distribution, increasing the risk of crop failure and food vulnerability in some regions. Socio-economic factors such as poverty levels, limited access to nutritious food, and inadequate public services further complicate food security conditions in this region. These conditions indicate that food security issues in North Aceh Regency are complex and multidimensional, thus requiring a systematic and objective approach to identify patterns of food security across regions based on measurable indicators. Furthermore, an interactive visualization medium is needed to present analysis results in a way that is easily understood by stakeholders in decision-making processes.

Data mining is a data analysis process that utilizes statistical techniques, artificial intelligence, and machine learning to identify patterns and extract meaningful knowledge from large-scale datasets [6]. Data mining includes the processes of data collection, data analysis, and information extraction to discover meaningful patterns in data [7]. One commonly used technique in data mining is clustering, which is a method used to group data based on similarity of characteristics without requiring predefined labels [8]. In addition to K-Medoids, another commonly used clustering method is K-Means due to its simplicity and computational efficiency. However, this method has a major weakness, as it is sensitive to outliers and extreme values because it uses the mean (centroid) as the center of the cluster. Furthermore, the clustering results in K-Means are highly dependent on the initial selection of centroids, which may lead to different clustering outcomes in each iteration. This condition makes K-Means less stable, especially when applied to data with high variation or uneven distribution [9].

Among various clustering algorithms, K-Medoids is known to be more robust against noise and outliers, making it more stable compared to methods such as K-Means [10]. K-Means Clustering is a non-hierarchical clustering method that aims to group data into a predetermined number of clusters based on the similarity of their characteristics. The K-Means process begins by determining the number of clusters, selecting initial centroids randomly, and then calculating the distance between each data point and the cluster centers using Euclidean distance.

In addition to K-Means, there is the K-Medoids algorithm, which uses medoids—actual data objects—as the centers of clusters instead of mean values. K-Medoids is more robust to outliers because it relies on real data representations. The clustering process in K-Medoids is performed by assigning data points based on their proximity to the selected medoids, resulting in more stable clustering outcomes compared to K-Means [11]. These findings indicate that the K-Medoids method has good capability in grouping data based on certain characteristics, making it relevant for food security analysis.

The evaluation of clustering quality in this study is carried out using the Davies-Bouldin Index (DBI), which is one of the internal evaluation metrics used to assess the quality of clustering results. This method measures the level of similarity within a cluster as well as the differences between clusters by considering the distance between cluster centers and the dispersion of data within clusters. A lower DBI value indicates better cluster separation and a more optimal structure. Thus, the use of DBI in this study aims to ensure that the clustering results can represent the data conditions more accurately and optimally [12].

Several previous studies have shown that the K-Medoids algorithm is effective in clustering agricultural and regional data. Research conducted by Hoki Dewangga et al. resulted in two main clusters in agricultural data grouping in Bengkulu Province, namely clusters with high and low crop yield potential [13]. Furthermore, research by Fadiah Irine Dwiana and Saiful Bahri identified two optimal clusters in food crop production, consisting of regions with high and low production levels [14]. In addition, research by Safruddin et al. produced five clusters based on corn production levels, ranging from very high to very low categories, which can be used as a basis for policy-making in improving regional food production [15]. However, most of these studies primarily focus on clustering regions based on agricultural production indicators and provide limited interpretation of the results. They do not comprehensively examine food security using multidimensional indicators that incorporate social, economic, and infrastructure aspects at the regional level, particularly in North Aceh Regency.

This gap highlights the need for clustering analysis that not only produces regional groupings but also provides meaningful interpretation to support policy-making and decision-making processes. Therefore, this study aims to cluster regions in North Aceh Regency based on food security levels using the K-Medoids algorithm, evaluate the quality of the clustering results, present the findings through web-based visualization, and provide policy recommendations based on the clustering outcomes. The results are expected to help local governments prioritize food insecurity interventions in a more targeted, data-driven manner, supporting effective policy planning and implementation, as well as improving understanding of regional disparities.

II. METHODOLOGY

This study applies the K-Medoids clustering algorithm to group regions based on food security indicators. The overall research process is illustrated in the flowchart shown in Figure 1.

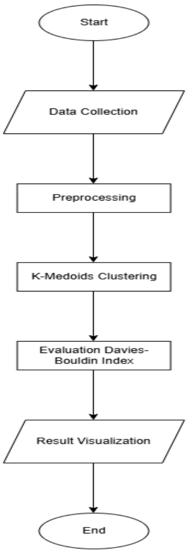

Figure 1. Research Flow of K-Medoids Clustering for Food Security Analysis

A. Data Collection

The data used in this study are food security data for North Aceh Regency in 2023 obtained from the Department of Agriculture and Food. This study was conducted from October to November 2025. North Aceh Regency consists of 27 sub-districts and 852 villages. The dataset in this study consists of nine variables, namely:

1. Agricultural land area (hectares)
2. Total area (hectares)
3. Number of food supply facilities and infrastructure
4. Number of population with the lowest welfare status
5. Villages without adequate transportation access
6. Number of households without access to clean water
7. Number of health workers
8. Total population
9. Number of households

These variables are used as the basis for constructing food security indicators representing the dimensions of food availability, food access, and food utilization.

B. Data Preprocessing

Before the clustering process, the data undergo a preprocessing stage to ensure data quality. This stage aims to prepare the data for analysis through data cleaning and dataset organization [16]. The preprocessing process includes calculating indicator ratios and data normalization.

The calculation of indicators in this study refers to the Food Security and Vulnerability Atlas (FSVA) method used by the Department of Agriculture and Food. Therefore, the indicators are not determined subjectively but are based on established standards for food security analysis. The indicators used in this study include:

1. Ratio of agricultural land area to total population
2. Ratio of food supply facilities and infrastructure to the number of households
3. Ratio of population with the lowest welfare status to total population
4. Villages without adequate transportation access via land, water, or air
5. Ratio of households without access to clean water to total households
6. Ratio of population per health worker relative to population density

The indicator of villages without adequate transportation access reflects the level of accessibility in a region. According to the Food Security and Vulnerability Atlas (FSVA) guidelines, this indicator is a condition-based variable expressed as an ordinal scale rather than a ratio. The scale ranges from 1 to 4, where: (1) accessible throughout the year, (2) generally accessible but may become inaccessible under certain conditions, such as infrastructure damage or temporary disruptions, (3) inaccessible during the rainy season, and (4) not accessible throughout the year.

For the indicator of population per health worker relative to population density, a composite ratio is applied. First, the ratio of population per health worker is calculated to represent the workload of each health worker. Then, this value is adjusted by population density to better reflect regional conditions, allowing differences between densely and sparsely populated areas to be considered.

Before normalization, the dataset is examined to ensure that there are no missing values or inconsistencies. Based on the results, no missing values are found, so no imputation process is required. Outlier handling is not specifically applied because, after converting the data into ratio form, the influence of extreme values becomes more controlled.

Next, data normalization is performed using the *Min-Max Normalization* method to transform values into a range between 0 and 1. This process aims to standardize the scale of attributes and prevent certain variables from dominating distance calculations [17]. The normalization formulas used are:

1. For positive indicators (higher values indicate better conditions):

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

2. For negative indicators (higher values indicate worse conditions):

$$x' = \frac{x_{max} - x}{x_{max} - x_{min}} \quad (2)$$

Where:

x' = normalized value

x = original value

x_{min} = minimum value of the attribute

x_{max} = maximum value of the attribute

Positive indicators include the ratio of agricultural land area to total population and the ratio of food supply facilities and infrastructure to the number of households. Negative indicators include the ratio of population with the lowest welfare status to total population, transportation access, the ratio of households without access to clean water, and the ratio of population per health worker relative to population density.

The classification of positive and negative indicators in this study follows the Food Security and Vulnerability Atlas (FSVA) guidelines used by the Department of Agriculture and Food of North Aceh Regency. Positive indicators represent conditions where higher values indicate better food security, while negative indicators represent conditions where higher values indicate greater vulnerability. After normalization, all indicators have a consistent interpretation, where higher values indicate better conditions.

The data processing in this study was conducted using Microsoft Windows 11 as the operating system, Google Chrome as the web browser, Visual Studio Code as the text editor, Laragon as the local server, and MySQL as the database system, with implementation using the PHP programming language. This setup ensures that the data processing can be reproduced and consistently applied in the clustering process.

C. K-Medoids Clustering

The clustering process in this study was carried out using the K-Medoids algorithm, a partitioning method in data mining used to group data based on the similarity of characteristics. This algorithm uses a medoid as the center of each cluster, which is the data object with the minimum total distance to other members within the cluster [18].

The determination of the number of clusters (k = 6) in this study is based on domain considerations referring to the official classification established by the Department of Agriculture and Food of North Aceh Regency. The institution defines six categories of food security levels, namely highly food insecure (C1), food insecure (C2), moderately food insecure (C3), moderately food secure (C4), food secure (C5), and highly food secure (C6). This approach ensures that the clustering results are interpretable and relevant for policy-making purposes.

Although alternative values of k could be evaluated, this study prioritizes a domain-based approach aligned with the official classification provided by the Department of Agriculture and Food. Therefore, experimental comparison with multiple k values is not conducted, as the objective is not to determine the optimal k mathematically but to ensure interpretability and policy relevance.

To support this selection, the clustering quality was evaluated using the Davies–Bouldin Index (DBI) and Silhouette Coefficient (SC) to assess the balance between intra-cluster compactness and inter-cluster separation. The evaluation results indicate that the clustering structure is acceptable. Therefore, the chosen number of clusters is considered appropriate for representing the variation in food security conditions in the study area. The steps of the K-Medoids algorithm in this study are as follows [19]:

1. Determine the number of clusters (k = 6)
2. Randomly select the initial medoids from the normalized dataset
3. Calculate the distance between each data point and the medoids using Euclidean Distance

   $$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{3}$$

   Where:

   $d(x, y)$ = distance between data point x and medoid y

   x_i = i-th attribute of data point x

   y_i = i-th attribute of medoid y

   n = number of attributes
4. Assign each data point to the nearest medoid based on the minimum distance.
5. Select candidate objects within each cluster as potential new medoids.
6. Recalculate the distance between each data point and the candidate medoids.
7. Compute the total deviation (S) between the current medoids and the candidate medoids.
8. If $S < 0$, replace the current medoids with the candidate medoids.
9. Repeat steps 3–8 until the medoids no longer change and the final clusters are formed.

In this study, the algorithm converged after six iterations, as indicated by no further changes in medoids and minimal reduction in total cost. This indicates that the clustering structure had stabilized, and additional iterations did not significantly improve the objective function. Therefore, the sixth iteration was considered sufficient to represent the final configuration.

To ensure the stability of the clustering results, the clustering process was repeated multiple times with different initial medoids, and the result with the lowest total cost was selected.

D. Evaluation Davies-Bouldin Index (DBI)

The Davies–Bouldin Index (DBI) is a cluster evaluation metric used to measure the quality of separation between clusters and the compactness within clusters. A lower DBI value indicates better clustering performance [20]. The steps for calculating DBI are as follows:

1. Calculate the intra-cluster dispersion (S_i), defined as the average distance between data points and the cluster centroid
2. Calculate the inter-cluster distance (d_{ij})
3. Compute the similarity ratio between clusters (R_{ij}) using the following formula:

$$R_{ij} = \frac{S_i + S_j}{d_{ij}} \quad (4)$$

Where:

R_{ij}: ratio value between cluster i and j

d_{ij}: distance between the centroids of cluster i and j

S_i, S_j: average distance of data points to the centroid in each cluster

4. Determine the maximum R_{ij} value for each cluster.
5. Calculate the DBI value as the average of these maximum values:

$$DBI = \frac{1}{k} \sum_{i=1}^{k} \max(R_{ij}),\ i \neq j \quad (5)$$

Where:

k: number of clusters

R_{ij}: ratio value between clusters

The DBI value is used in this study to evaluate the quality of the clustering results obtained.

E. Silhouette Coefficient (SC)

The Silhouette Coefficient is an evaluation metric used to assess the quality of clustering results by measuring both the cohesion within clusters and the separation between clusters. The Silhouette Coefficient ranges from −1 to 1. A value close to 1 indicates that an object is well matched to its own cluster, while a value close to −1 indicates that the object is more appropriately assigned to another cluster. A value around 0 suggests that the object lies between two clusters [21]. The steps for calculating the Silhouette Coefficient are as follows:

1. Compute the average distance between a data point and all other points within the same cluster, denoted as $a(i)$

$$a(i) = \frac{1}{|A|-1} \Sigma_{j \in A, j \neq i}\, d(i,j) \quad (6)$$

Where:

A = number of data points in the same cluster

2. Compute the average distance between the data point and all points in another cluster, and take the minimum value as $b(i)$

$$d(i,C) = \frac{1}{|C|} \Sigma_{j \in C}\, d(i,j) \quad (7)$$

Where:

C = number of data points in another cluster

3. Calculate the Silhouette Coefficient for each data point using the following formula:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (8)$$

The value of $s(i)$ ranges from −1 to 1, with the following interpretations:

$s(i)$ close to 1 indicates that the data point is well clustered within its group

$s(i)$ close to 0 indicates that the data point lies between two clusters

$s(i)$ close to −1 indicates that the data point may better assigned to another cluster.

The average Silhouette Coefficient value is used in this study to evaluate the overall quality of the clustering results.

III. RESULT AND DISCUSSION

This section presents the research findings, including clustering results, visualization, evaluation, and their interpretation for policy implications.

A. Data Description

The dataset consists of 852 villages in North Aceh Regency and includes nine variables representing key aspects of food security, as summarized in Table I.

TABLE I
INPUT VARIABLES FOR FOOD SECURITY ANALYSIS

Code	Variable	Unit
V1	Agricultural land area	ha
V2	Regional area	ha
V3	Food facilities and infrastructure	unit
V4	Low-income population	person
V5	Regional accessibility	ordinal
V6	Households without clean water access	household
V7	Number of health workers	person
V8	Total population	person
V9	Number of households	household

These variables form the basis for constructing food security indicators used in the clustering process. Table II presents a subset of the dataset to illustrate its structure.

TABLE II
RESEARCH DATASET

No	Sub-district	Village	V1	V2	V3	V4	V5	V6	V7	V8	V9
1	Sawang	Riseh Tunong	341.00	2825.65	7	768	1	110	2	517	721
2	Sawang	Gunci	154.00	2175.63	7	304	1	64	5	2548	678
3	Sawang	Kubu	65.87	495.90	1	0	1	83	3	681	203
4	Sawang	Blang Cut	27.12	1356.59	1	42	1	52	2	613	139
5	Sawang	Riseh Teungoh	50.56	765.57	2	0	1	53	2	3523	133
…	…	…	…	…	…	…	…	…	...	…	…
852	Dewantara	Bluka Teubai	3.00	130.01	22	0	1	130	11	1841	505

B. *Data Preprocessing*

The preprocessing stage was conducted through ratio calculation and data normalization using the Min-Max method.

1. Ratio Calculation

The indicators used in this study are derived from the Food Security and Vulnerability Atlas (FSVA) framework provided by the Department of Agriculture and Food of North Aceh Regency and are presented in Table 3.

Code	Variable	Unit
X1	Ratio of agricultural land area to total population	Ratio
X2	Ratio of food supply facilities and infrastructure to the number of households	Ratio
X3	Ratio of population with the lowest welfare level to total population	Ratio
X4	Village without adequate transportation access via land, water, or air	Nominal
X5	Ratio of households without access to clean water to the total number of households	Ratio
X6	Ratio of village population per health worker to population density	Ratio

The ratio scale indicates that the variables are expressed as proportional values, while the nominal scale represents categorical variables. The ratio indicators are defined as follows:

$$X_1 = \frac{V_1}{V_8} \tag{9}$$

$$X_2 = \frac{V_3}{V_9} \tag{10}$$

$$X_3 = \frac{V_4}{V_8} \tag{11}$$

$$X_4 = V_5 \tag{12}$$

$$X_5 = \frac{V_6}{V_9} \tag{13}$$

$$X_6 = \left(\frac{v_8}{v_7}\right) : \left(\frac{v_8}{v_2}\right) \tag{14}$$

As an example, the ratio calculation for Riseh Tunong Village, Sawang District, is as follows:

$$X_1 = \frac{341.00}{517} = 0.659$$

$$X_2 = \frac{7}{721} = 0.009$$

$$X_3 = \frac{768}{517} = 1.485$$

$$X_4 = 1$$

$$X_5 = \frac{110}{721} = 0.152$$

$$X_6 = \left(\frac{517}{2}\right) : \left(\frac{517}{2825.65}\right) = 14.128$$

Additionally, a subset of the ratio calculation results is presented in Table 4. The same calculation formula was consistently applied to all villages in the dataset.

All values are presented with three decimal places for consistency therefore, integer values are written as 1.000.

TABLE IV
RATIO CALCULATION RESULTS

No	Sub-district	Village	X1	X2	X3	X4	X5	X6
1	Sawang	Riseh Tunong	0.659	0.009	1.485	1.000	0.152	14.128
2	Sawang	Gunci	0.604	0.010	0.119	1.000	0.094	4.351
3	Sawang	Kubu	0.967	0.004	0.000	1.000	0.408	1.653
4	Sawang	Blang Cut	0.044	0.007	0.068	1.000	0.374	6.782
5	Sawang	Riseh Teungoh	0.015	0.015	0.000	1.000	0.398	3.827
…	…	…	…	…	…	…	…	…
852	Dewantara	Bluka Teubai	0.001	0.043	0.000	1.000	0.257	0.118

2. Data Normalization

The classification of indicators into positive and negative indicators follows the definition provided in Section 2B. The minimum and maximum values of each indicator are presented in Table 5.

TABLE V
MINIMUM AND MAXIMUM VALUES OF INDICATORS

Indicator	Minimum Value	Maximum Value
X1	0.000	7.462
X2	0.000	0.833
X3	0.000	1.485
X4	1.000	2.000
X5	0.000	2.638
X6	0.000	71.042

1. For positive indicators, the normalization is calculated as follows:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

2. For negative indicators, the normalization is calculated as follows:

$$x' = \frac{x_{max} - x}{x_{max} - x_{min}} \quad (2)$$

These values are obtained from the entire dataset consisting of 852 villages. As an example, the normalization for Riseh Tunong Village, Sawang District, is carried out using the minimum and maximum values from all villages in the dataset. The same calculation procedure is consistently applied to all indicators based on the values presented in Table 5, as follows:

$$X_1 = \frac{0.659 - 0.000}{7.462 - 0.000} = 0.088$$

$$X_2 = \frac{0.009 - 0.000}{0.833 - 0.000} = 0.011$$

$$X_3 = \frac{1.485 - 1.485}{1.485 - 0.000} = 0.000$$

$$X_4 = \frac{2.000 - 1.000}{2.000 - 1.000} = 1.000$$

$$X_5 = \frac{2.638 - 0.152}{2.638 - 0.000} = 0.942$$

$$X_6 = \frac{71.042 - 14.128}{71.042 - 0.000} = 0.801$$

Normalized values range from 0 to 1, where values closer to 1 indicate better conditions and values closer to 0 indicate worse conditions. The same normalization procedure is applied to all villages to ensure consistency. A subset of the normalized data is presented in Table 6.

TABLE VI
NORMALIZED DATA

No	Sub-district	Village	X1	X2	X3	X4	X5	X6
1	Sawang	Riseh Tunong	0.088	0.011	0.000	1.000	0.942	0.801
2	Sawang	Gunci	0.008	0.012	0.919	1.000	0.964	0.938
3	Sawang	Kubu	0.012	0.005	1.000	1.000	0.845	0.976
4	Sawang	Blang Cut	0.005	0.008	0.953	1.000	0.858	0.904
5	Sawang	Riseh Teungoh	0.002	0.018	1.000	1.000	0.848	0.946
…	…	…	…	…	…	…	…	…
852	Dewantara	Bluka Teubai	0.000	0.052	1.000	1.000	0.902	0.998

C. K-Medoids Clustering

1. Initial Medoids Selection

The initial medoids are determined as temporary cluster centers in the K-Medoids clustering process. The medoids are selected randomly from the normalized dataset. These medoids are then used as reference points in calculating distances between data points. The initial medoids used in this study are presented in Table 7.

TABLE VII
INITIAL MEDOIDS

No	Sub-district	Village	X1	X2	X3	X4	X5	X6
1	Sawang	Riseh Tunong	0.088	0.011	0.000	1.000	0.942	0.801
2	Sawang	Gunci	0.008	0.012	0.919	1.000	0.964	0.938
3	Sawang	Kubu	0.012	0.005	1.000	1.000	0.845	0.976
4	Sawang	Blang Cut	0.005	0.008	0.953	1.000	0.858	0.904
5	Sawang	Riseh Teungoh	0.002	0.018	1.000	1.000	0.848	0.946
6	Sawang	Riseh Baroh	0.011	0.011	0.953	1.000	0.980	0.942

2. Distance Calculation Using Euclidean Distance

Distance calculation is performed to measure the proximity between each data point and the medoids. In this study, Euclidean Distance is used to compute the distance between a data point and the first medoid.

$$d_1 = \sqrt{\begin{matrix}(0.088 - 0.088)^2 + (0.011 - 0.011)^2 + \\ (0.000 - 0.000)^2 + (1.000 - 1.000)^2 + \\ (0.942 - 0.942)^2 + (0.801 - 0.801)^2\end{matrix}}$$

The distance between a data point and the second medoid is calculated as follows:

$$d_2 = \sqrt{\begin{matrix}(0.008 - 0.088)^2 + (0.012 - 0.011)^2 + \\ (0.919 - 0.000)^2 + (1.000 - 1.000)^2 + \\ (0.964 - 0.942)^2 + (0.938 - 0.801)^2\end{matrix}}$$

The same process is applied to all data points for each medoid. The resulting distances are used to assign each data point to the nearest cluster based on the minimum distance. Table 8 presents a subset of the distance calculation results from the first iteration.

TABLE VIII
K-MEDOIDS CALCULATION RESULTS (ITERATION 1)

Iteration 1							
No	**C1**	**C2**	**C3**	**C4**	**C5**	**C6**	**Distance**
1	0	0.933	1.022	0.966	1.018	0.967	0
2	0.933	0	0.148	0.116	0.140	0.037	0
3	1.022	0.148	0	0.087	0.034	0.147	0
4	0.966	0.116	0.867	0	0.063	0.128	0
5	1.018	0.140	0.034	0.063	0	0.140	0
…	…	…	…	…	…	…	
852	1.024	0.124	0.077	0.121	0.082	0.114	0.077
Total Distance							**109.588**

3. Determination of New Medoids
After the clustering step, the medoids are updated by selecting candidate medoids from the existing data. This process aims to obtain more optimal cluster centers. The initial medoids used in this study are presented in Table 9.

TABLE IX
UPDATE MEDOIDS

No	Sub-district	Village	X1	X2	X3	X4	X5	X6
1	Sawang	Lhok Cut	0.019	0.017	1.000	1.000	0.858	0.920
2	Sawang	Sawang	0.004	0.021	1.000	1.000	0.929	0.989
3	Sawang	Blang Teurakan	0.005	0.007	1.000	1.000	0.884	0.982
4	Sawang	Jurong	0.009	0.010	1.000	1.000	0.873	0.969
5	Sawang	Blang Manyak	0.005	0.026	0.904	1.000	0.969	0.848
6	Sawang	Lhok Jok	0.017	0.043	1.000	1.000	0.813	0.903

4. Recalculation Using Euclidean Distance
After obtaining the updated medoids, the distance between each data point and the medoids is recalculated in the second iteration using the Euclidean Distance method. This step aims to update cluster assignments based on the new medoids.

The distance between a data point and the first cluster center (medoid) is calculated as follows:

$$d_1 = \sqrt{\begin{matrix}(0.088 - 0.019)^2 + (0.011 - 0.017)^2 + \\ (0.000 - 1.000)^2 + (1.000 - 1.000)^2 + \\ (0.942 - 0.858)^2 + (0.801 - 0.920)^2\end{matrix}}$$

The distance between a data point and the second cluster center (medoid) is calculated as follows:

$$d_2 = \sqrt{\begin{matrix}(0.008 - 0.019)^2 + (0.012 - 0.017)^2 + \\ (0.919 - 1.000)^2 + (1.000 - 1.000)^2 + \\ (0.964 - 0.858)^2 + (0.938 - 0.902)^2\end{matrix}}$$

The same calculation process is then applied to all data points for each medoid. The resulting distances are used to reassign each data point to the nearest cluster based on the minimum distance value. This process produces an updated clustering structure for the second iteration, as presented in Table X.

TABLE X
K-MEDOIDS CALCULATION RESULTS (ITERATION 2)

Iteration 2							
No	**X1**	**X2**	**X3**	**X4**	**X5**	**X6**	**Distance**
1	1.013	1.021	1.021	1.019	0.909	1.016	0.909
2	0.134	0.101	0.121	0.124	0.092	0.177	0.092
3	0.058	0.087	0.040	0.030	0.203	0.088	0.030

4	0.051	0.120	0.094	0.081	0.135	0.074	0.051
5	0.031	0.091	0.051	0.035	0.182	0.063	0.031
…	…	…	…	…	…	…	
852	0.096	0.042	0.050	0.058	0.191	0.131	0.042
Total Distance							**106.903**

5. Calculation of Total Deviation (S)
After obtaining the updated medoids, the distance between each data point and the medoids is recalculated in the second iteration using the Euclidean Distance method. This step is essential to update cluster memberships and to ensure that each data point is assigned to the most appropriate cluster based on the newly selected medoids. By recalculating the distances, the clustering structure is refined and improved in each iteration. The total deviation is then computed to evaluate whether the clustering process should continue or stop. The deviation value (S) is calculated as the difference between the new total distance and the previous total distance, as shown below:
S = Total Distance New – Total Distance Old
= 106.903 – 109.588
= -2.685
Since $S < 0$, this indicates that the total distance has decreased, meaning that the clustering result has improved.

Therefore, the medoid update process is continued by selecting new medoids from non-medoid objects. This iterative process ensures that the algorithm searches for a more optimal set of medoids that minimizes the total distance within clusters. The iteration then proceeds using the same sequence of steps, including distance calculation, cluster assignment, and medoid updating, until convergence is achieved. Convergence occurs when there is no significant change in medoids or when the total distance no longer decreases substantially between iterations. Based on the results obtained from the system, the clustering process continues until the sixth iteration. However, the lowest total cost is achieved in the fifth iteration, with a value of 89.201. Therefore, the fifth iteration is selected as the final clustering result, as it represents the most optimal clustering configuration obtained in this study. For clarity and brevity, manual calculations are presented only up to the second iteration, while subsequent iterations follow the same computational procedure. The final clustering results are presented in Table XI.

TABLE XI
FINAL RESULTS OF K-MEDOIDS CLUSTERING

No	X1	X2	X3	X4	X5	X6	Distance	Cluster	Label
1	1.018	1.022	1.029	1.008	1.036	1.002	1.002	C6	Highly food secure
2	0.133	0.124	0.200	0.100	0.298	0.081	0.081	C6	Highly food secure
3	0.028	0.071	0.070	0.062	0.272	0.125	0.028	C1	Highly food insecure
4	0.093	0.121	0.113	0.096	0.275	0.126	0.093	C1	Highly food insecure
5	0.047	0.082	0.072	0.068	0.262	0.120	0.047	C1	Highly food insecure
…	…	…	…	…	…	…	…	…	…
852	0.055	0.030	0.119	0.045	0.244	0.089	0.030	C2	Food insecure

To further support the interpretation, the average values of each indicator for every cluster are presented in Table XII. These values provide a clearer overview of the distribution of indicator characteristics within each cluster, allowing a more objective and data-driven understanding of the differences in food security conditions across regions.

TABLE XII
AVERAGE NORMALIZED INDICATOR VALUES PER CLUSTER

Cluster	X1	X2	X3	X4	X5	X6
C1	0.014	0.035	0.985	0.984	0.852	0.985
C2	0.013	0.082	0.991	0.956	0.907	0.992
C3	0.016	0.063	0.981	0.983	0.756	0.983
C4	0.011	0.026	0.971	0.976	0.919	0.987
C5	0.053	0.363	0.989	0.959	0.947	0.951
C6	0.016	0.064	0.982	0.991	0.991	0.978

Table XII presents the average normalized indicator values (X1–X6) for each cluster, which serve as the basis for interpreting the characteristics of food security levels. The results show clear variations across clusters that support the classification of food security conditions. Cluster C1, categorized as highly food insecure, exhibits low values in positive indicators such as agricultural land ratio (X1 = 0.014) and food infrastructure (X2 = 0.035), indicating limited production capacity and infrastructure availability. These conditions suggest that villages within this cluster have constrained resources to support adequate food production and distribution, which may contribute to higher levels of vulnerability. At the same time, it shows relatively high values in negative indicators, including welfare vulnerability (X3 = 0.985) and limited transportation access (X4 = 0.984).

These conditions reflect unfavorable socio-economic and accessibility conditions. In contrast, cluster C6, categorized as highly food secure, demonstrates more favorable conditions across most indicators, particularly in access to clean water (X5 = 0.991) and health worker availability (X6 = 0.978), which represent strong support systems for food security. Intermediate clusters (C2–C5) show gradual variations in indicator values. For example, cluster C5 has the highest values in agricultural land ratio (X1 = 0.053) and food infrastructure (X2 = 0.363), indicating strong production and infrastructure capacity, although some vulnerability indicators remain relatively high. Meanwhile, cluster C3 and C4 represent moderate conditions with mixed indicator values, reflecting transitional levels between vulnerable and secure categories. Overall, these results confirm that the cluster labeling is consistent with the underlying data distribution and is supported by the average indicator values within each cluster.

TABLE XIII
CHARACTERISTICS OF EACH CLUSTER

Cluster	Characteristics	Category
C1	Low land ratio, high poverty, poor water & transport access	Highly food insecure
C2	Moderate vulnerability indicators	Food insecure
C3	Slightly improved conditions	Moderately food insecure
C4	Better infrastructure and welfare	Moderately food secure
C5	Good conditions	Food secure
C6	Best conditions across indicators	Highly food secure

To provide empirical justification for the cluster labeling, each cluster is analyzed based on the dominant characteristics of its indicators. These characteristics are derived from the observed patterns of indicator values within each cluster. The labeling follows the classification established by the Department of Agriculture and Food. Cluster C1 is characterized by high vulnerability indicators, such as poor access to clean water, limited transportation, and low welfare levels, which supports its classification as highly food insecure. In contrast, cluster C6 shows favorable conditions across most indicators, indicating a high level of food security. Other clusters represent intermediate conditions between vulnerable and secure categories. To strengthen the interpretation of cluster labels, each cluster is characterized based on the average values of its indicators. For instance, clusters categorized as highly food insecure tend to exhibit lower agricultural land ratios, limited infrastructure availability, and reduced access to essential services. In contrast, highly food secure clusters show higher values across these indicators.

Based on the clustering results, it was found that cluster C1 (highly food insecure) consists of 63 villages, cluster C2 (food insecure) consists of 92 villages, cluster C3 (moderately food insecure) consists of 179 villages, cluster C4 (moderately food secure) consists of 42 villages, cluster C5 (food secure) consists of 49 villages, and cluster C6 (highly food secure) consists of 427 villages.

These clustering results are subsequently used as a basis for determining policy interventions for each group of villages. Villages in clusters C1 and C2 are prioritized for three types of interventions, clusters C3 and C4 receive two types of interventions, cluster C5 receives one type of intervention, while cluster C6 is focused on maintaining its current level of food security.

The clustering performance is evaluated using the Davies–Bouldin Index (DBI) and Silhouette Coefficient (SC). The DBI value of 3.085 indicates a moderate level of clustering quality in terms of intra-cluster compactness, suggesting that data points within each cluster are relatively close to their respective medoids, although some variation within clusters may still exist. Meanwhile, the SC value of 17.75% suggests that the separation between clusters is relatively weak, indicating that the boundaries between clusters are not clearly defined. These results indicate that, although the clustering structure achieves a reasonable level of compactness, the separation between clusters is still limited, suggesting the presence of overlapping characteristics among clusters.

D. *Clustering Results Visualization*

Visualization is used to provide a clear overview of the distribution of food security across regions.

Figure 2. Bar chart showing the number of villages in each food security cluster (C1–C6) in North Aceh Regency

Figure 2 shows the distribution of villages into six food security clusters in North Aceh Regency. The data are presented in a bar chart, where the X-axis represents the cluster categories (C1–C6) and the Y-axis indicates the number of villages. Each bar represents the number of villages in each cluster category. Cluster C1 (highly food insecure) includes 63 villages, C2 (food insecure) includes 92 villages, C3 (moderately food insecure) includes 179 villages, C4 (moderately food secure) includes 42 villages, C5 (food secure) includes 49 villages, and C6 (highly food secure) includes 427 villages. Cluster C6 contains the highest number of villages, indicating that overall food security conditions in North Aceh Regency are relatively good. However, a number of villages remain in the highly food insecure (C1) and food insecure (C2) categories, reflecting disparities across regions. These disparities may be associated with variations in food availability, access, and utilization, as well as differences in resources, infrastructure, and socio-economic conditions. The visualization not only illustrates the distribution of clusters but also highlights the variation in key indicators across regions. Differences in agricultural land ratio, infrastructure availability, and access to basic services contribute significantly to the observed clustering patterns.

A more detailed distribution at the sub-district level is presented in Figure 3, which provides deeper insight into cluster distribution at the local level. In the developed system, users can select a specific sub-district to view food security conditions in more detail. As an example, Sawang Sub-district is selected to illustrate the food security conditions of villages within the area.

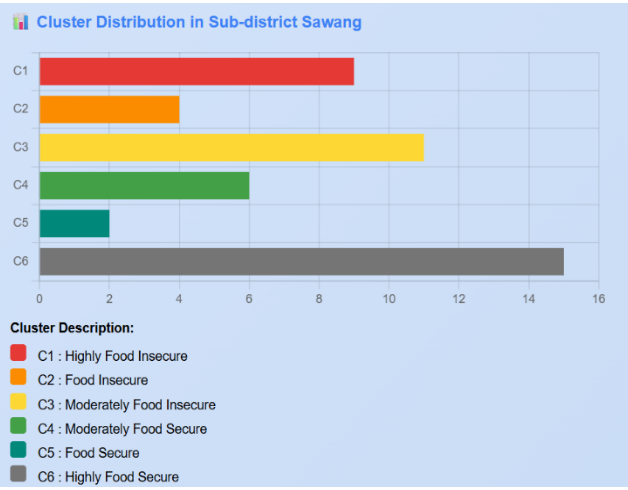

Figure 3. Bar chart showing the number of villages in each food security cluster (C1–C6) in Sawang Sub-district

Figure 3 presents the distribution of villages based on food security clusters at the sub-district level, specifically in Sawang Sub-district. The bar chart shows the number of villages in each cluster category, where the X-axis represents the clusters (C1–C6) and the Y-axis shows the number of villages. The results indicate that Sawang Sub-district has 9 villages in cluster C1 (highly food insecure), 4 villages in C2 (food insecure), 11 villages in C3 (moderately food insecure), 6 villages in C4 (moderately food secure), 2 villages in C5 (food secure), and 15 villages in C6 (highly food secure). This shows that most villages fall into the highly food secure category. These findings indicate that food security conditions in Sawang Sub-district are relatively good and tend to be stable, although some villages still experience vulnerability. Clusters C1 and C2 are identified as priority targets for policy intervention, particularly in improving food availability, access, and utilization, as well as strengthening infrastructure and socio-economic conditions.

The distribution shown in Figure 3 is further visualized spatially in Figure 4 to illustrate the geographical distribution of each cluster category in Sawang Sub-district.

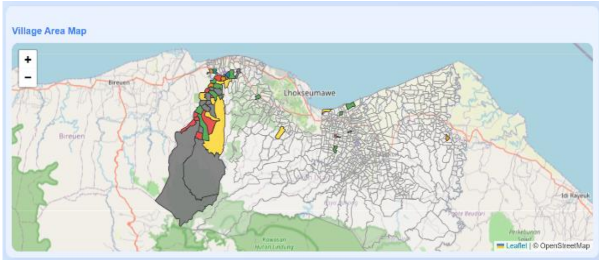

Figure 4. Thematic map illustrating the spatial distribution of food security clusters across villages in Sawang Sub-district

Figure 4 illustrates the spatial distribution of village food security clusters in Sawang Sub-district using a thematic map. Each color on the map represents a different food security cluster category, and each village is displayed according to its corresponding cluster, allowing the visualization of geographical patterns in food security levels. The results show that food security is unevenly distributed, with clusters ranging from highly food insecure (C1) to highly food secure (C6) across different areas. Even within the same sub-district, variations in cluster categories can be observed, indicating that food security conditions are influenced by local factors related to food availability, access, and utilization, such as differences in resources, infrastructure, and socio-economic conditions. This spatial pattern highlights the importance of area-based analysis and targeted policy interventions to reduce disparities and improve overall food security.

E. Discussion

The clustering results show that villages are divided into six groups of food security levels with distinct characteristics based on the indicator values. A more detailed analysis reveals that the differences between clusters are primarily influenced by specific indicators rather than general conditions.

Cluster C1 (highly food insecure) is characterized by low values of positive indicators, particularly the agricultural land ratio (X1) and food facilities ratio (X2), indicating limited food availability and infrastructure. In addition, this cluster shows high values in vulnerability indicators such as the proportion of low-welfare population (X3), limited access to clean water (X5), and poor transportation accessibility (X4). These conditions indicate that villages in this cluster face significant challenges in food access and utilization.

Cluster C2 (food insecure) shows similar patterns to C1 but with slightly improved indicator values, suggesting moderate vulnerability. Clusters C3 and C4 represent transitional conditions, where improvements can be observed in certain indicators such as infrastructure (X2) and access factors (X4 and X5), although some weaknesses still persist.

In contrast, clusters C5 and C6 are characterized by higher values in positive indicators and lower values in vulnerability indicators. In particular, cluster C6 (highly food secure) shows favorable conditions across most indicators, including better agricultural capacity (X1), improved infrastructure (X2), and better access to clean water and health services (X5 and X6), indicating strong food availability, access, and utilization.

These findings suggest that food security in North Aceh Regency is strongly influenced by the balance between availability, access, and utilization indicators. The dominance of cluster C6 indicates generally favorable conditions; however, the presence of clusters C1 and C2 highlights persistent disparities that require targeted intervention.

The results of this study are consistent with previous research that applied clustering methods in agricultural and regional data analysis. For example, studies by Dewangga et al. [13] and Dwiana and Bahri [14] identified clusters representing high and low production regions, while Safruddin et al. [15] produced multiple clusters based on production levels. These findings confirm that clustering methods are effective in grouping regions based on similar characteristics.

However, unlike previous studies that mainly focused on production-based indicators, this study incorporates a more comprehensive set of food security indicators, including availability (X1), infrastructure (X2), welfare (X3), accessibility (X4), clean water access (X5), and health services (X6). This allows for a more detailed interpretation of food security conditions, not only in terms of production but also access and utilization.

Furthermore, this study provides a more in-depth interpretation by linking cluster results with specific indicator patterns, enabling clearer identification of the factors contributing to food insecurity and food security at the village level.

F. Policy Implications

The policy recommendations in this study are derived from the key indicators influencing each cluster, including agricultural land availability (X1), food infrastructure (X2), welfare level (X3), transportation access (X4), clean water access (X5), and health services (X6).

The classification of villages based on food security levels can serve as a basis for more targeted policy-making. Villages in clusters C1 and C2 should be prioritized for intervention, particularly in improving agricultural land utilization (X1), enhancing food facilities and infrastructure (X2), increasing access to clean water (X5), improving transportation infrastructure (X4), and strengthening healthcare services (X6), as well as addressing welfare issues (X3).

Clusters C3 and C4 require policies focused on improving weaker indicators, especially in terms of infrastructure development, accessibility, and basic service provision. Meanwhile, clusters C5 and C6 should focus on maintaining and enhancing existing conditions to ensure long-term sustainability of food security.

With this clustering approach, local governments can more effectively determine policy priorities and allocate resources according to regional conditions. Moreover, this data-driven approach supports more targeted and sustainable development planning to improve food security.

IV. Conclusion

Based on the results of data processing using the K-Medoids algorithm, the following conclusions can be drawn:

1. The implementation of the K-Medoids algorithm successfully grouped 852 villages into six clusters: highly food insecure (C1) with 63 villages, food insecure (C2) with 92 villages, moderately food insecure (C3) with 179 villages, moderately food secure (C4) with 42 villages, food secure (C5) with 49 villages, and highly food secure (C6) with 427 villages. This result indicates that most villages fall into the highly food secure category, although a considerable number of villages are still classified as food insecure, reflecting disparities in food security conditions across regions, which can serve as a basis for prioritizing policy interventions in vulnerable areas.
2. The evaluation using the Silhouette Coefficient (SC) yielded a value of 17.75%, indicating that the separation between clusters is relatively weak. This suggests that cluster boundaries are not clearly defined and that overlapping characteristics exist among clusters.
3. The evaluation using the Davies–Bouldin Index (DBI) resulted in a value of 3.085, indicating a moderate level of cluster compactness, where data points within clusters are relatively close to their respective medoids, although some variation within clusters still exists.

References

[1] Republik Indonesia, *Undang-Undang Republik Indonesia Nomor 18 Tahun 2012 Tentang Pangan*, vol. 1. Indonesia, 2012, pp. 1–7.

[2] N. Z. Tiftazani, M. Rifkie, S. Noerzaman, and C. G. Hermawan, “Evaluasi Dan Perkembangan Ketahanan Pangan Indonesia Tahun 2023 – 2025 Berdasarkan Kerangka Empat Pilar FAO,” vol. 2, pp. 851–873, 2025.

[3] Kementerian Pertanian, “Analisis Komoditas Pangan Strategis Tahun 2023,” p. 87, 2023.

[4] Pemerintah Republik Indonesia, *Peraturan Pemerintah Republik Indonesia Nomor 17 Tahun 2015 Tentang Ketahanan Pangan dan Gizi*. 2015, pp. 1–63.

[5] M. Viona, D. G. Katanging, and M. Candra, “Ekonomi Politik Ketahanan Pangan di Indonesia : Peran Negara Dalam Menghadapi

Krisis Pangan," vol. 02, no. 6, pp. 502–508, 2025.
[6] A. A. Baskara, N. M. Piranti, and M. F. Romdendine, "Framework Data Mining : Sebuah Survei," vol. 9, no. 3, pp. 4886–4895, 2025.
[7] A. Safrin, S. Khan, M. Fatekurohman, and Y. S. Dewi, "Perbandingan Algoritma K-Medoids Dan K-Means Dalam Pengelompokan Kecamatan Berdasarkan Produksi Padi Dan Palawija Di Jember," vol. 2, no. 2, pp. 67–75, 2023.
[8] M. A. Bui and A. Bahtiar, "Implementasi Metode Algoritma K-Means Clustering Untuk Mengelompokkan Transaksi Penjualan Barang Di Toko Arino," vol. 8, no. 2, pp. 1451–1456, 2024.
[9] D. Anggraini and M. S. Hasibuan, "Initial Centroid Pada Algoritma K-Means Dan K-Medoids," vol. 07, no. 03, pp. 487–500, 2025.
[10] M. D. Salman *et al.*, "Perbandingan Kinerja Algoritma Clustering K-Means dan K-Medoids dalam Pengelompokan Sekolah di Provinsi Riau Berdasarkan Ketersediaan Sarana dan Prasarana: Comparison of K-Means and K-Medoids Clustering Algorithm Performance in Grouping Schools in Riau Pr," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 5, no. 3, pp. 797–806, 2025.
[11] R. Khairuna and A. Razi, "Analisis perbandingan algoritma k-means dan k-medoids untuk klusterisasi teks ulasan pada aplikasi Cookpad," vol. 10, no. 2, pp. 445–458, 2025.
[12] Y. Hasan, U. B. Darma, K. Medan, and S. Utara, "Pengukuran Silhouette Score Dan Davies- Bouldin Index Pada Hasil Cluster K-Means Dan DBSCAN," vol. 12, no. 3, 2024.
[13] H. Dewangga, K. Khairil, and I. Y. Beti, "Penerapan Metode K-Medoid Dalam Pengelompokkan Data Pertanian Di Dinas Tanaman Pangan Hortikultura Dan Perkebunan Provinsi Bengkulu," *J. Media Infotama*, vol. 20, no. 2, pp. 501–509, 2024.
[14] F. I. Dwiana and S. Bahri, "Implementasi Algoritma K-Medoids pada Klasterisasi Produksi Jenis Tanaman Pangan di Provinsi Jawa Timur," *J. Mat. Integr.*, vol. 20, no. 1, pp. 25–34, 2024.
[15] S. Safruddin, J. W. Sitopu, A. A. Manurung, I. Satria, and A. Wanto, "Pengelompokkan Produksi Tanaman Jagung di Sumatera Utara Menggunakan Algoritma K-Medoids," *J. Media Inform. Budidarma*, vol. 7, no. 1, p. 484, 2023.
[16] J. M. Polgan *et al.*, "Penggunaan Teknik Data Mining untuk Analisis Perilaku Pengguna pada Media Sosial," vol. 13, pp. 1074–1078, 2024.
[17] P. P. Allorerung, A. Erna, and M. Bagussahrir, "Analisis Performa Normalisasi Data untuk Klasifikasi K-Nearest Neighbor pada Dataset Penyakit," vol. 9, no. 3, pp. 178–191, 2024.
[18] D. Z. Putri, R. De Pani, S. Ginting, L. Efrizoni, and R. Rahmaddeni, "Penerapan Algoritma K-Medoids dalam Menganalisis Pola Pelanggan untuk Strategi Pemasaran," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 12, no. 3, 2025.
[19] A. Hoerunnisa, G. Dwilestari, F. Dikananda, H. Sunana, and D. Pratama, "Komparasi Algoritma K-Means Dan K-Medoids Dalam Analisis Pengelompokan Daerah Rawan Kriminalitas Di Indonesia," 2024.
[20] R. Ishak and Amiruddin, "Clustering Prestasi Akademik Lulusan Menggunakan Metode K-Means," vol. 6, 2024.
[21] A. Akram *et al.*, "Mini-Batch K-Means Clustering Untuk Pengelompokan Kemandirian Daerah Di Sulawesi Selatan," pp. 235–244, 2024.