Journal of Applied Food Technology

Home page: https://ejournal2.undip.ac.id/index.php/jaft

Multivariate Exploration of Food Security in the Sulampua Region: Identification of Clusters and Dominant Dimensions of Food Security

Wawan Saputra*, Naufalia Alfiryal, I Putu Gde Inov Bagus Prasetya, Anwar Fitrianti, Kevin Alifviansyah

Department of Statistics and Data Science, School of Science, Mathematics, and Informatics, IPB University, Dramaga Raya Street, IPB Dramaga Campus, Bogor, Indonesia

*Corresponding author (w2nwawan@ipb.ac.id)

Abstract

Food security is a strategic issue closely related to economic development, community welfare, and the achievement of sustainable development goals. The Food Security Index (FSI) is an important instrument for measuring food security conditions at the provincial and district/city levels. However, FSI performance in Indonesia still shows regional disparities, particularly in Sulawesi, Maluku, and Papua (Sulampua), which tend to have low scores. This study aims to explore patterns of food security and vulnerability in Sulampua through multivariate analysis and regional clustering using K-Means and K-Medoids (PAM) methods. The analysis begins with Principal Component Analysis (PCA) to reduce the dimensionality of FSI indicators and identify dominant factors contributing to data variation. The PCA results show that the first three components explain more than 77% of the variance, with dominant factors including poverty, food expenditure, basic infrastructure access, as well as health and nutrition indicators. The clustering analysis produces two main groups: cluster 1, which includes the majority of districts/cities in Sulawesi and Maluku with relatively better food security, and cluster 2, consisting of 16 districts/cities in Papua with significant food insecurity. Cluster validity evaluation indicates that the K-Medoids method performs better than K-Means, being more robust to outliers and producing more consistent cluster separation. This study contributes to the literature by providing multivariate visual exploration and regional classification based on FSI indicators, which can serve as a basis for formulating more targeted food security policies in the Sulampua region.

Article information:
Received: 17 October 2025
Accepted: 13 December 2025
Available online: 16 December 2025

Keywords:
Food security
PCA
K-Means
K-Medoids
Sulampua

doi: 10.17728/jaft.29754

Introduction

The agricultural sector is one of the most important sectors in Indonesia, as the majority of its population is employed in and relies on this sector as a primary source of livelihood (Azhar and Sulianto, 2023). In addition to serving as the main provider of food, the agricultural sector also contributes significantly to the Gross Domestic Product (GDP), generates foreign exchange, and absorbs a large share of the workforce (Khairiyakh et al., 2015). In other words, the resilience of the agricultural sector is closely related to national economic stability, community welfare, and sustainable development.

Food security is a strategic issue that has attracted both national and global attention. According to Law No. 18 of 2012 on food, national food security refers to a condition in which food is sufficiently available in terms of quantity, quality, safety, diversity, nutrition, equity, and affordability. This is in line with the definition provided by the Food and Agriculture Organization (FAO), which emphasizes four main dimensions of food security: availability, access, utilization, and stability (FAO, 2013). In the context of sustainable development, food security represents one of the key targets of the Sustainable Development Goals (SDGs), particularly Goal 2 (Zero Hunger).

In Indonesia, one of the instruments commonly used to measure food security is the National Food Security Index (Indeks Ketahanan Pangan, FSI), developed by the National Food Agency with reference to the Global Food Security Index (GFSI). The FSI measures the level of food security at the provincial and district/city levels by considering three aspects: food availability, accessibility or affordability, and food utilization. The FSI serves as a comprehensive measure of food security and functions as a reference for policymaking, identifying regional disparities, and setting development priorities (Badan Pangan Nasional, 2022).

However, food security achievements across Indonesia remain uneven. This is reflected by the low FSI scores of 70 out of 416 districts/cities, with most low-scoring regions located in Eastern Indonesia, particularly

Sulawesi, Maluku, and Papua (collectively referred to as Sulampua) (Badan Pangan Nasional, 2022).

The low FSI scores in the Sulampua region are influenced by complex and multi-dimensional factors reflected in its underlying indicators. From the aspect of food availability, issues such as geographic isolation, limited productive land, and distribution barriers caused by poor transportation infrastructure are key challenges. In terms of food access, high poverty rates, higher food prices compared to western regions, and limited market access exacerbate food insecurity. From the perspective of food utilization, malnutrition, limited access to clean water, and inadequate sanitation remain persistent problems in several areas. This complexity demonstrates that food security challenges in the Sulampua region are not only economic in nature but also social, geographic, and infrastructural (Badan Pangan Nasional, 2022).

Given these conditions, an in-depth study is necessary to better understand the patterns of food security and vulnerability across Sulampua. One suitable method for this purpose is cluster analysis, which groups data with similar (homogeneous) characteristics into the same cluster, while dissimilar (heterogeneous) data are assigned to different clusters (FAO, 2013).

There are several types of cluster analysis, two of the most commonly used being k-means and k-medoids. K-means is a simple clustering method that measures similarity based on the distance between objects and the cluster centroid (mean). Its advantages include ease of implementation, efficiency with large datasets, and straightforward interpretation of clustering results (Zhang et al., 2020). Similar to k-means, k-medoids is also a partition-based clustering method; however, it uses actual data points as cluster centers, known as medoids (Prihandoko et al., 2017). This makes k-medoids more robust to outliers and noise compared to k-means (Arora et al., 2016).

Several previous studies have applied similar approaches. For instance, Ula et al. (2023) applied the Fuzzy C-Means method to cluster food-insecure regions in North Aceh District based on Food Security Index (FSI) indicators. Their study produced three clusters, highly insecure, insecure, and moderately insecure and employed the Borda method to determine priority areas for intervention. Similarly, Hafsari, Hijrah, Wijayanti, and Kurniawan (2024) implemented the k-means algorithm to cluster 514 districts/cities across Indonesia based on food security indicators, identifying four clusters categorized as very high, high, moderate, and low.

These studies indicate that cluster analysis has been widely used in food security research. However, previous studies generally employed only a single clustering method or focused on national or localized scales. Few have specifically compared k-means and k-medoids methods to cluster regions based on Food Security Index indicators, particularly in the Sulawesi, Maluku, and Papua (Sulampua) regions. Moreover, earlier studies rarely emphasized multivariate visual exploration, which can provide a more informative understanding of food security and vulnerability patterns across regions. Therefore, this study aims to present multivariate data exploration and visualization to illustrate the patterns of food security and vulnerability, classify districts/cities in Sulawesi, Maluku, and Papua (Sulampua) based on FSI indicators, and identify the key factors that contribute most significantly to these conditions.

Unlike previous studies that primarily applied a single clustering method or focused on national-level analysis, this study offers a comparative evaluation of K-Means and K-Medoids clustering methods specifically for the Sulawesi, Maluku, and Papua (Sulampua) regions. In addition, this study emphasizes multivariate visual exploration through Principal Component Analysis (PCA) to reveal dominant dimensions of food security and regional vulnerability patterns. This approach contributes to the existing literature by providing a more detailed and region-specific understanding of food security disparities in Eastern Indonesia.

Materials and Methods

Materials

This study utilizes data from the Food Security Index (Indeks Ketahanan Pangan, FSI) indicators of the Sulampua region for the period 2018–2024, obtained from the official website of the National Food Agency (Badan Pangan Nasional, Bapanas). The dataset consists of 139 observations and 13 variables. The analysis was conducted using administrative identifiers (Province and District/City) and nine core Food Security Index (FSI) indicators, namely National Consumption per Capita Ratio (NCPR), Poverty Percentage, Food Expenditure Percentage, Percentage Without Electricity, Percentage Without Clean Water, Average Years of Schooling for Women, Health Worker Ratio, Life Expectancy, and Stunting Percentage. The Food Security Index (FSI) score was used as an outcome reference for interpretation and cluster characterization. Other variables such as year and population were excluded from the multivariate analysis as they serve only as descriptive metadata.

Method

To achieve the research objectives, several analytical stages were carried out systematically, beginning from data collection to the interpretation of cluster results. The stages of this study are described as follows:

1. Data Exploration
   Descriptive analysis was conducted to understand the characteristics of the data, including measures of central tendency and dispersion. Univariate and bivariate visualizations were used to identify initial patterns and relationships among variables.
2. Data Pre-processing
   - All variables were standardized using the *z-score* method to ensure comparability and prevent bias resulting from differences in measurement scales.
   - Prior to analysis, the dataset was examined for missing values. The results indicated that no missing values were present in the selected Food Security Index indicators. Therefore, no imputation procedure was required.
3. Dimensionality Reduction using Principal Component Analysis (PCA)
   PCA was employed to identify the variables that

contributed most to the overall data variation and to determine the dominant dimensions influencing food security and vulnerability. PCA was performed on the standardized Food Security Index (FSI) indicators. PCA was applied to reduce dimensionality, identify dominant dimensions of food security, and mitigate multicollinearity among indicators.

4. Determination of the Optimal Number of Clusters using the Elbow Method
The optimal number of clusters was determined using the Elbow Method by identifying the inflection point in the plot of the number of clusters versus the within-cluster sum of squares (WCSS).
5. Cluster Analysis using K-Means and K-Medoids Methods
Cluster analysis was performed using the dominant dimensions obtained from Principal Component Analysis (PCA) of the nine key Food Security Index (FSI) indicators defined by the National Food Agency. The retained principal components were used as inputs in the clustering analysis to ensure that the classification of regions was based on the most informative and non-redundant dimensions derived from the original indicators. The k-means and k-medoids algorithms were employed to identify regional food security patterns. The FSI score itself was not included in the PCA and clustering process, as it is a composite index derived from the same indicators and was used only for result interpretation and cluster characterization.
6. Cluster Evaluation
The clustering results were evaluated using several validation metrics, including the Dunn Index, Silhouette Score, and Calinski-Harabasz Index.
7. Cluster Interpretation and Policy Implications
The identified clusters were interpreted to gain insights into the regional patterns of food security and vulnerability in Sulampua, providing evidence-based recommendations for targeted policy interventions.

Manhattan Distance

In cluster analysis, the choice of distance measurement method is highly relevant, as it directly affects the grouping results. Several types of distance measures can be used; one of the most common is the Manhattan distance, also known as the city block distance. This measure calculates the distance between an object and the cluster center by summing the absolute differences across all variables, as defined in the following equation:

$$d_{cb}(x_i, c_k) = \sum_{j=1}^{p} \left| x_{ij} - c_{kj} \right|, i = 1,2,3, \dots, n \text{ and } k = 1,2,3, \dots, k \quad (1)$$

In this equation, $d_{cb}(x_i, c_k)$ represents the Manhattan distance between the *i* object and the *k* cluster center, x_{ij} denotes the value of the *i* object for the *j* variable, and c_{kj} indicates the centroid value of the *k* cluster for the *j* variable. Here, *n* represents the total number of objects analyzed, p denotes the number of variables, and K refers to the number of clusters specified.

K-Means

The K-Means algorithm is a simple clustering method that uses distance measures to assess the degree of similarity between objects. The basic principle of this method is that the smaller the Manhattan distance between objects, the greater their similarity. The process is performed iteratively to produce a number of clusters corresponding to the predetermined value of K. One of the main advantages of K-Means is its relatively fast and efficient implementation on large datasets, as well as the ease of interpreting the clustering results (Zhang et al., 2020).

The algorithm begins by specifying the number of clusters (K) and selecting the initial cluster centers (centroids). Each object is then assigned to the nearest centroid, and the centroid positions are updated based on the mean position of the cluster members. This process continues iteratively until there are no further changes in the cluster assignments (Prihandoko et al., 2017).

K-Medoids

K-Medoids, also known as PAM (Partitioning Around Medoids), is a partition-based clustering method that selects actual data points as the cluster centers (medoids) (Nahdliyah et al., 2019). Unlike K-Means, which uses the mean value as the cluster center, K-Medoids chooses medoids that are true members of the dataset, making it more representative and robust against noise and outliers. Although its computational time is generally longer than that of K-Means, the difference is not significant for small-sized datasets. This method is particularly suitable for clustering data obtained from multiple measurement points.

The main stages of the K-Medoids algorithm can be summarized as follows (Shang et al., 2022):

1. Determine the number of clusters (K) and randomly select several data points as the initial medoids.
2. Compute the manhattan distance between each data point and all medoids, then assign each data point to the nearest medoid.
3. For each cluster, identify the data point with the smallest total distance to other members and set it as the new medoid.
4. Repeat the process until the medoids remain stable or the maximum number of iterations is reached.

Principal Component Analysis

Principal Component Analysis (PCA) is a dimensionality reduction technique designed to transform high-dimensional data into a lower-dimensional representation while retaining as much of the original variance as possible (Savira, 2023). This method converts a large number of correlated variables into a smaller set of new, mutually independent variables known as principal components.

Mathematically, PCA performs a linear transformation on correlated variables to form a new set of uncorrelated variables. Each principal component is the projection result of this transformation and is ordered according to the amount of variance it explains. The first principal component represents the greatest variation in the data, while the second component captures the remaining variation not explained by the first, and so on (Firliana et al., 2015).

Elbow Method

The Elbow method is used to determine the optimal number of clusters (k) in cluster analysis. This method plays an important role in identifying the most appropriate number of clusters when the number of clusters is not known in advance. The optimal value of k is indicated by a sharp bend or the most significant decrease between two consecutive cluster numbers in the plotted graph. This comparison is obtained by calculating the Sum of Squared Errors (SSE) for each value of k.

The SSE measures the total within-cluster variation by summing the squared distances between each observation and the centroid of its assigned cluster. As the number of clusters increases, the SSE value tends to decrease because data points are assigned to clusters with closer centroids. The Elbow method identifies the point at which adding more clusters does not result in a substantial reduction in SSE, indicating an optimal trade-off between model complexity and clustering performance. The SSE formulation used in the Elbow method can be expressed as follows (Dewi and Pramita, 2019):

$$SSE = \sum_{K=1}^{K} \sum_{x_i} |x_i - c_k| \tag{2}$$

With,
K = the *c* cluster
x_i = distance of the *i*data object
c_k = center of the *k* cluster

Evaluation Metrics

In this study, several evaluation techniques are employed to assess the performance of the clustering analysis, namely:

1. Dunn Index
   The Dunn Index is used to measure the relationship between cluster compactness and separation. A higher Dunn Index value indicates a better clustering result. The Dunn Index can be calculated using the following equation (Dun, 1974):
   $$D(u) = \min_{1 \le i \le k} \left\{ \min_{1 \le j \le k, j \ne 1} \left\{ \frac{(\delta(X_{i,} X_j)}{\max\limits_{1 \le c \le k} \{\Delta(X_c)\}} \right\} \right\} \tag{3}$$
   $\delta(X_i, X_j)$ represents the distance between clusters, generally computed from the centroids of clusters C_i and C_j, while $\Delta(Xc)$ denotes the intra-cluster distance within cluster *Xc* (Dunn, 1974).

2. Silhouette Score
   The Silhouette Score is an evaluation metric in cluster analysis used to assess how well a clustering algorithm forms distinct groups. This metric measures how similar each data point is to other points within the same cluster compared to points in other clusters. The Silhouette Score ranges from -1 to 1, where a value closer to 1 indicates that the clusters are well separated and the data points are appropriately grouped. The formula for the Silhouette Score is given as follows (Kossakov et al., 2024):
   $$S(i) = \frac{b_{(i)} - a_{(i)}}{\max(a_{(i)}, b_{(i)})} \tag{4}$$
   Here, *b(i)* represents the smallest average distance between data point *i* and all points in a different cluster (i.e., the nearest neighboring cluster to which *i* does not belong). Meanwhile, *a(i)* denotes the average distance between data point *i* and all other points within the same cluster, for a dataset *X* with *n* data points. The Silhouette Score for data point *i* is then calculated as follows (Kossakov et al., 2024):
   $$Silhoutte\ Score = \frac{1}{\mathrm{n}} \sum_{i=1}^{n} S(i) \tag{5}$$

3. Calinski-Harabasz Index
   The Calinski–Harabasz Index is used to evaluate the quality of clustering results by comparing the Sum of Squares Between Clusters (SSB), which represents the degree of separation between clusters, and the Sum of Squares Within Clusters (SSW), which reflects the compactness within clusters. The ratio between these two quantities is multiplied by a normalization factor derived from the ratio of the total number of observations to the number of clusters formed minus one. This index, also known as the Variance Ratio Criterion (VRC), is useful for determining the optimal number of clusters. A higher Calinski–Harabasz Index value indicates better clustering performance, meaning the clusters are more compact internally and well separated from one another. The formula for calculating the Calinski–Harabasz Index is expressed as follows (Aselnino and Wijayanto, 2023):
   $$CH\,(k) = \frac{B(k)/(k-1)}{W(k)/(n-k)} \tag{6}$$
   where $B(k)$ and $W(k)$ denote the between-cluster and within-cluster variances, respectively. The variable k represents the number of clusters (k > 1), while n refers to the total number of observations or elements (Aselnino and Wijayanto, 2023).

Results and Discussion

Statistic Descriptive Analysis

Descriptive analysis was conducted to provide an initial overview of the variation among the indicators that compose the Food Security Index (FSI) in the Sulampua region. The results indicate considerable differences in the distribution across indicators, particularly within the dimensions of food access and utilization.

Figure 1 presents the boxplots of the FSI indicators, illustrating the distribution patterns, median values, and the presence of outliers for each indicator. The plot shows that the poverty rate, food expenditure, and stunting indicators exhibit relatively high variability compared to the others, indicating disparities in food security across regions. In contrast, the health worker

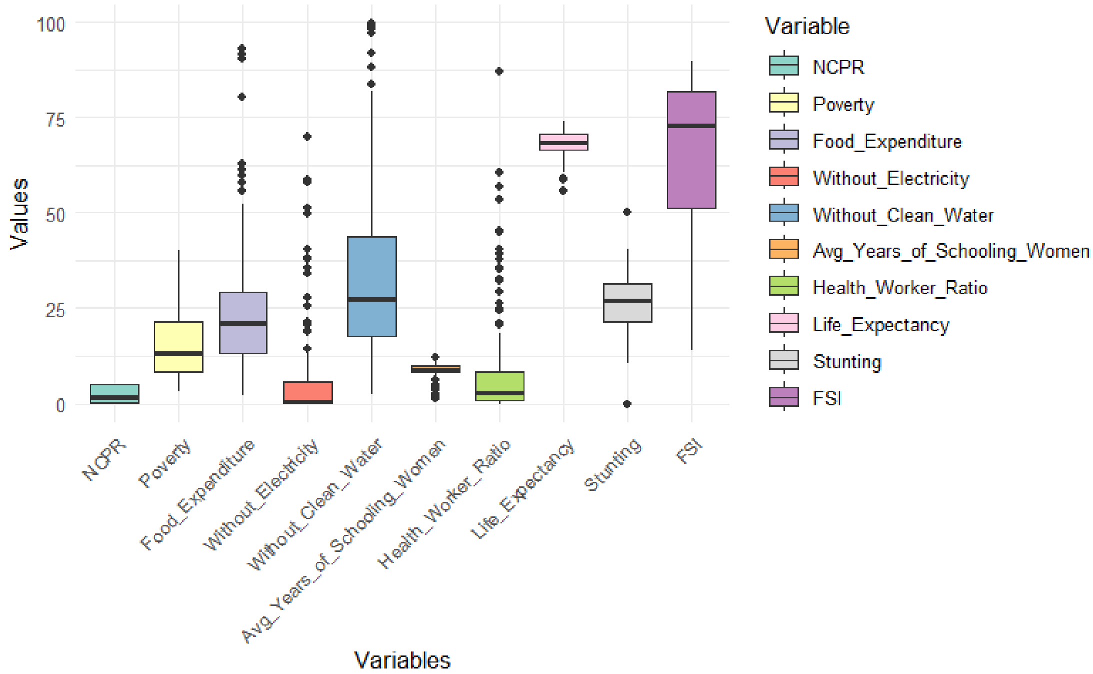

Figure 1. Boxplots of Food Security Index (FSI) Indicators for Sulampua

ratio displays a narrower spread, suggesting relatively smaller variations in healthcare access among districts and cities. This implies that, compared to economic and nutritional dimensions, access to basic health services is more evenly distributed and may play a less dominant role in explaining regional food security disparities.

Figure 2 illustrates the spatial distribution of the Food Security Index (FSI) scores across the Sulampua region. The thematic map shows that most districts and cities in Papua and Maluku tend to fall into the vulnerable category, represented by dark purple to blue colors indicating low FSI scores (below 50). Meanwhile, moderate food security levels are depicted by green to yellow colors, while high FSI scores are shown in orange to red colors.

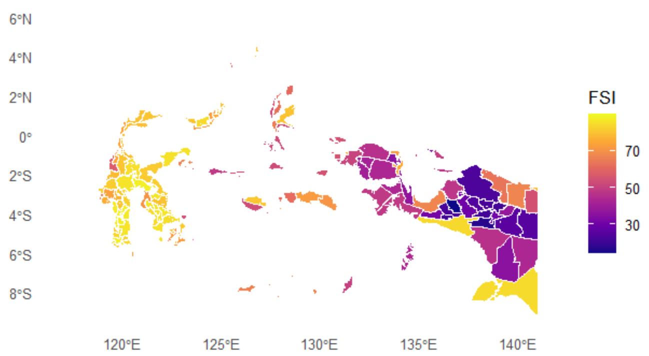

Figure 2. Food Security Index (FSI) Score Distribution in the Sulampua Region

The Sulawesi region exhibits a more heterogeneous pattern. Several districts and cities, particularly major urban centers such as Makassar, Kendari, and Manado, show relatively high FSI scores, reflecting better food security conditions. However, vulnerable areas are still present within Sulawesi, indicating intra-island disparities in food security levels.

Beyond the visual pattern, these spatial disparities reflect structural differences in socioeconomic conditions, geographic accessibility, and infrastructure development across regions. Papua and Maluku are characterized by challenging topography, dispersed settlements, and limited transportation and market access, which constrain food availability and distribution. In contrast, higher FSI scores in Sulawesi are associated with more developed infrastructure, stronger market integration, and better access to public services. These findings are consistent with food security theory, which emphasizes the role of access and utilization, alongside availability, as key determinants of regional food security.

The observed east–west gap within Sulampua also aligns with previous studies on food security in Eastern Indonesia, which report persistent vulnerabilities in Papua and Maluku due to structural and geographic disadvantages. Differences between clusters therefore imply not only variations in food consumption outcomes but also deeper socioeconomic and infrastructural inequalities that shape regional resilience and vulnerability. This spatial pattern provides a critical foundation for subsequent multivariate and clustering analyses aimed at identifying groups of regions with similar food security characteristics.

Visual Multivariate Exploration

Although descriptive analysis provides an overview of the distribution of individual food security indicators, it is insufficient to capture their interrelationships and the underlying multidimensional structure of food security. Therefore, multivariate exploration using Principal Component Analysis (PCA) was conducted to jointly analyze the indicators, reduce dimensionality, and address potential multicollinearity. PCA was further used to identify the dominant dimensions contributing to regional differences in food security. The number of principal components retained was determined using a scree plot based on the proportion of variance explained by each component.

Figure 3 presents the Scree Plot resulting from the Principal Component Analysis (PCA). It can be observed that the first principal component (PC1) explains approximately 56.1% of the total data variance, followed by PC2 with 12.5% and PC3 with 8.7%. Thus, the first three principal components together account for more than 77% of the total variation, which is sufficient to represent the food security patterns in the Sulampua region.

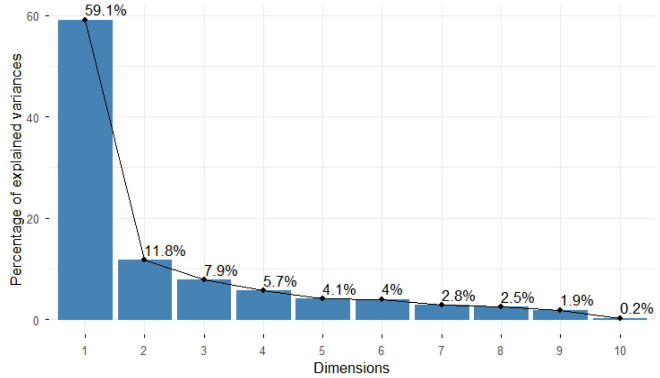

Figure 3. Scree Plot of Principal Component Analysis (PCA)

After determining the number of principal components, the next step is to identify the contribution of each indicator to the respective components. The PCA loading results, which show the coefficient values of each indicator on PC1, PC2, and PC3, are presented in Table 1.

Based on the results in Table 1, PC1 is dominated by indicators related to poverty (0.385), food expenditure (0.315), lack of electricity (0.368), and lack of clean water (0.385). The dominance of these indicators reflects the strong interconnection between economic deprivation and inadequate basic infrastructure, which jointly constrain food access and utilization. Regions with high poverty levels tend to experience limited access to electricity and clean water, increasing household vulnerability and reducing the ability to obtain, store, and utilize food effectively. Therefore, PC1 can be interpreted as a dimension of economic vulnerability and basic infrastructure access.

These dimensions play a critical role in shaping regional disparities in food security across Sulampua. High PC1 scores are predominantly observed in Papua and parts of Maluku, where challenging geography, dispersed settlements, limited transportation networks, and uneven infrastructure development restrict market access and public service delivery. In contrast, lower PC1 scores in many districts of Sulawesi reflect relatively better infrastructure, stronger economic integration, and improved access to basic services, contributing to higher food security outcomes observed in these regions.

Meanwhile, PC2 is primarily influenced by stunting (0.699) and women's years of schooling (0.226), representing the health and nutrition dimension of food security. This component highlights the role of human capital and nutritional status in shaping long-term food security outcomes beyond economic access alone. Furthermore, PC3 is characterized by contributions from women's years of schooling (0.378) and life expectancy (-0.292), which can be associated with the education and demographic dimension, reflecting differences in social development and population health across regions. Overall, the PCA results provide a multivariate explanation for the observed spatial and cluster-based food security patterns, linking structural socioeconomic and geographic conditions to regional food security outcomes.To clarify the relationships between indicators and principal components, a correlation heatmap was utilized. This visualization facilitates the identification of interrelated indicators and those that contribute most significantly to the formation of the principal dimensions.

Figure 4 presents the correlation heatmap between the FSI indicators and the retained principal components, providing a visual validation of the PCA loading results. The heatmap highlights the strength and direction of relationships between indicators and each component. Indicators related to poverty, food expenditure, lack of electricity, and lack of clean water exhibit strong positive correlations with PC1, while life expectancy and women's years of schooling show

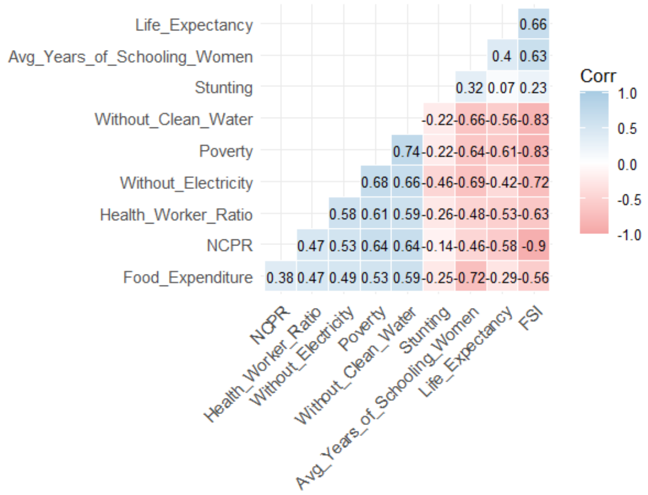

Figure 4. Correlation Heatmap of Food Security Index (FSI) Indicators

Table 1. Principal Component Loadings of Food Security Index (FSI) Indicators

Indicator	PC1	PC2	PC3
NCPR	0.327	0.322	0.18
Poverty	0.385	0.146	0.017
Food Expenditure	0.315	-0.24	-0.605
Without Electricity	0.368	-0.231	0.164
Without Clean Water	0.385	0.11	-0.101
Average Years of Schooling for Women	-0.362	0.226	0.378
Health Worker Ratio	0.333	0.059	0.178
Life Expectancy	-0.301	-0.457	-0.292
Stunting	-0.169	0.699	-0.552

negative correlations. This contrasting correlation pattern confirms that PC1 captures opposing conditions between socio-economic deprivation and human development across regions.

K-Means and K-Medoids Cluster Analysis

Cluster analysis was carried out using two approaches, namely K-Means and K-Medoids, based on both the original FSI indicators and the principal components derived from PCA. The optimal number of clusters was determined using the Elbow Method, taking into account the number of FSI categories reported by the National Food Agency. Simulations were performed with varying numbers of clusters, k =2,3,4,5, and 6.

Figure 5 presents the elbow plot used to determine the optimal number of clusters for the K-Means and K-Medoids methods. The fundamental principle of the Elbow Method is to identify the “elbow point” on the curve depicting the relationship between the number of clusters (k) and the Within-Cluster Sum of Squares (WCSS). As the number of clusters increases, the WCSS value decreases because the distance between cluster members and their centroids becomes smaller. However, beyond a certain point, the rate of decrease slows down, indicating that adding more clusters no longer significantly improves clustering quality.

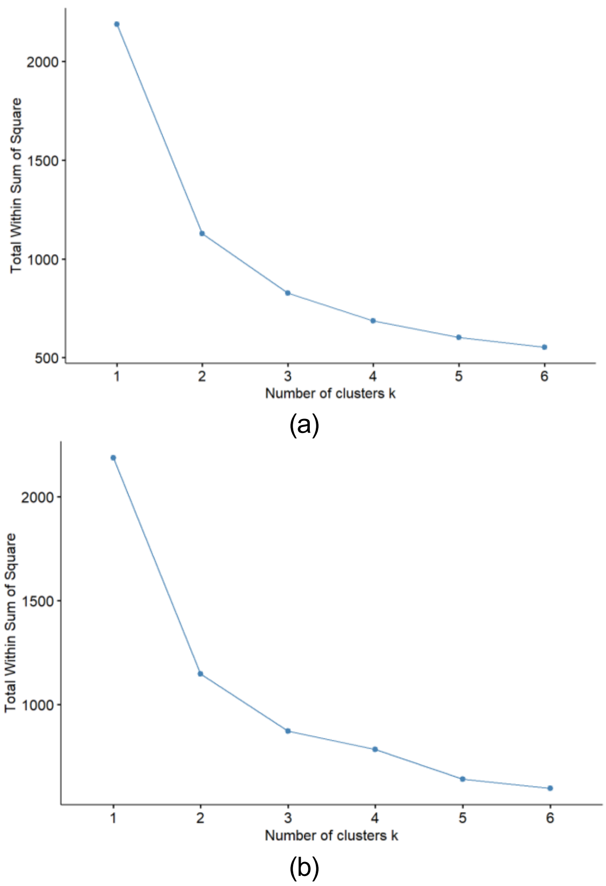

Figure 5. Elbow Plots for Determining the Optimal Number of Clusters Using (a) K-Means and (b) K-Medoids Methods

In this analysis, the elbow point is clearly observed at k = 2, suggesting that two clusters are optimal for representing the structure of food security data in the Sulampua region. With this number of clusters, a balance is achieved between clear group separation and efficient cluster representation. The selection of k = 2 also aligns with the objective of distinguishing between regions with relatively high food security and those that are more vulnerable. This finding provides a strong foundation for subsequent analyses using both K-Means and K-Medoids methods, where districts and cities are grouped into two primary clusters based on similarities in FSI indicator characteristics and PCA-derived components.

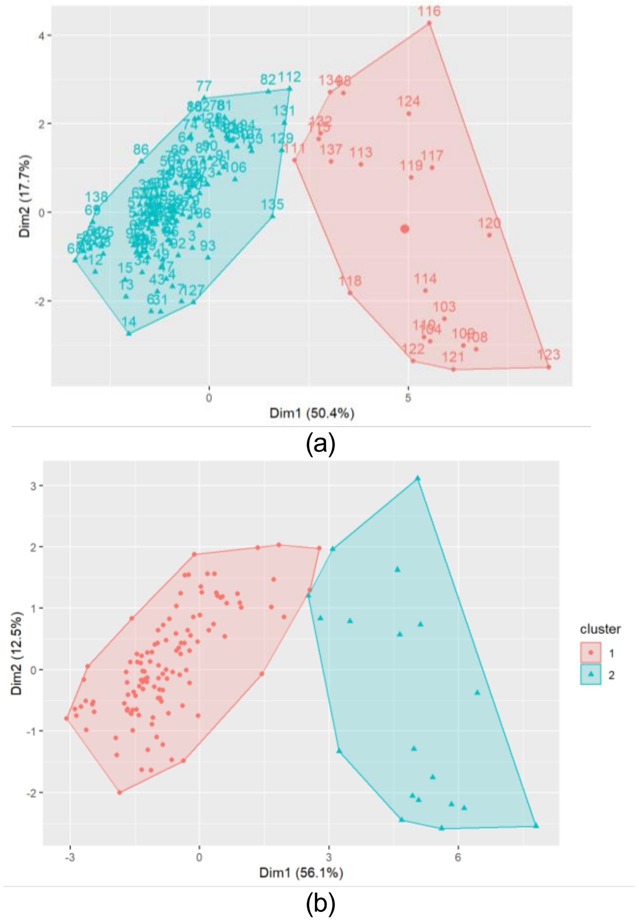

Figure 6. Cluster plots of Food Security Index (FSI) indicators generated using (a) K-Means and (b) K-Medoids algorithms.

Figure 6 displays the clustering results using the K-Means method (left panel) and the K-Medoids method (right panel) with the optimal number of clusters k = 2. The visualization is presented in a two-dimensional space derived from PCA projection, allowing for easier observation of group separation patterns. In the K-Means results, most districts and cities are concentrated in Cluster 1 (blue), while Cluster 2 (red) consists of a group of regions with distinctly different characteristics. Some points are located near the cluster boundaries, indicating that K-Means is relatively sensitive to the presence of outliers, as cluster centers are determined based on centroids.

In contrast, the K-Medoids clustering results show a more compact cluster distribution and clearer separation between groups. This aligns with the characteristics of the K-Medoids method, which uses actual data points (medoids) as cluster centers, making it more stable against outliers. Therefore, although both methods produce two main groups—regions with

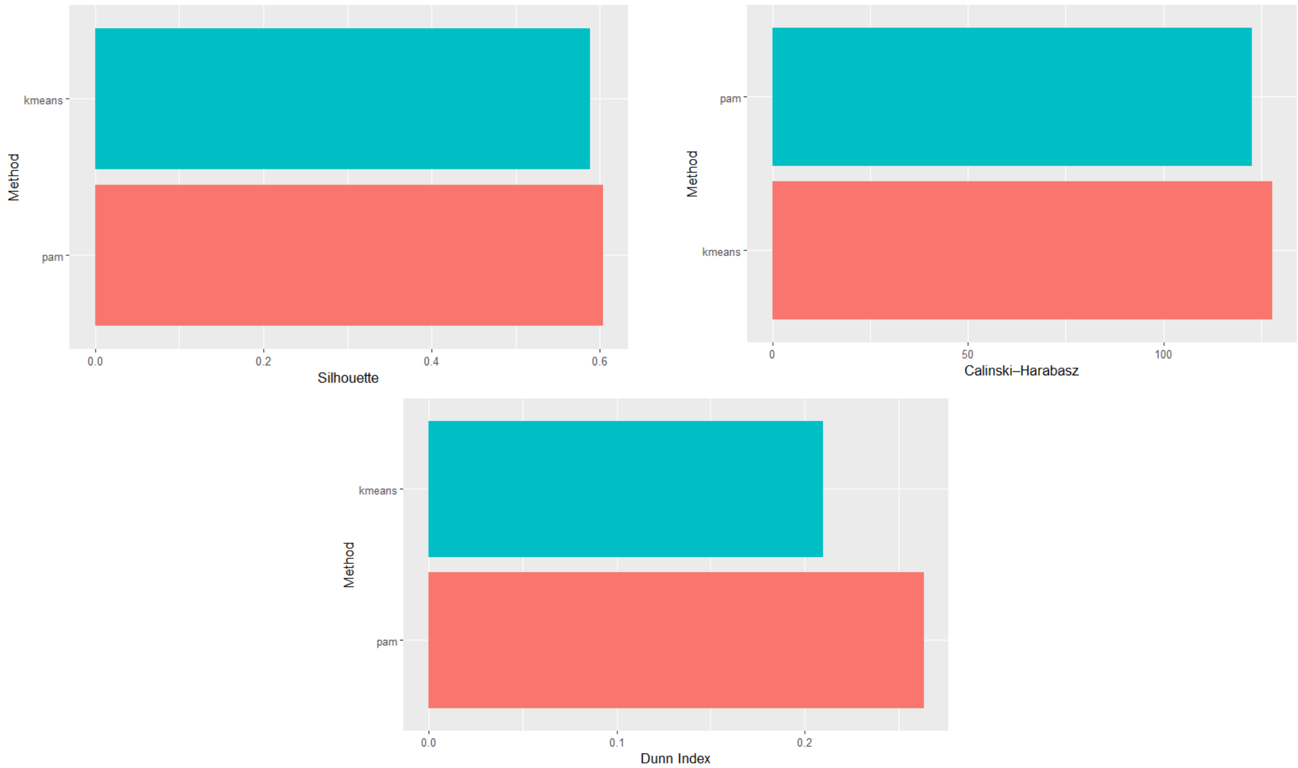

Figure 7. Comparative Evaluation of Clustering Performance between K-Means and K-Medoids Methods

relatively strong food security and those with higher vulnerability, the K-Medoids method is considered more consistent and robust, making it more suitable for food security data that may contain extreme values.

Cluster analysis performance was then evaluated to compare the clustering quality of both methods. Three evaluation metrics were used: Silhouette Score, Calinski–Harabasz Index, and Dunn Index. The evaluation results are presented in Figure 7.

Based on Figure 7, it can be observed that the silhouette and Dunn index values are higher for the K-Medoids method, indicating that it produces more compact clusters with clearer separation between groups. Meanwhile, the Calinski–Harabasz index is slightly higher for K-Means, suggesting relatively better internal cluster cohesion. Overall, although both methods yield reasonably good clustering results, K-Medoids tends to perform better due to its robustness against outliers and its ability to group extreme regions more consistently.

The main difference between the two methods lies in the stability of cluster membership. In K-Means, several extreme regions shift between clusters, whereas in K-Medoids, their placement remains more consistent. The number of cluster members obtained from both methods is illustrated in Figure 8.

Based on the clustering results using the K-Medoids method, two main groups were identified, with the distribution of members shown in Figure 8.

- Cluster 1 consists of 122 districts/cities (88.4%), and is characterized by high heterogeneity, as it includes regions with both very high and relatively low Food Security Index (FSI) scores. Examples of districts/cities with high FSI scores in this cluster include Kendari City (89.67), Ternate City (88.24), Makassar City (87.95), Sidenreng Rappang (87.92), and Soppeng (87.90). However, this cluster also contains areas with low FSI scores, such as Teluk Wondama (36.76), South Sorong (40.39), Maybrat (41.27), Teluk Bintuni (43.12), and Aru Islands (43.15).
- Cluster 2 comprises 16 districts/cities (11.6%), all located in Papua. This cluster is relatively homogeneous and predominantly represents regions with low food security. Some areas in this cluster with relatively higher FSI scores include Asmat (45.99), Yalimo (37.80), Pegunungan Arfak (35.53), Mappi (35.44), and Supiori (32.69). Meanwhile, districts such as Intan Jaya (14.14), Nduga (17.10), Lanny Jaya (21.33), Deiyai (21.35), and Central Mamberamo (21.51) exhibit extremely low FSI scores and represent the most severe food security vulnerability identified by the cluster analysis.

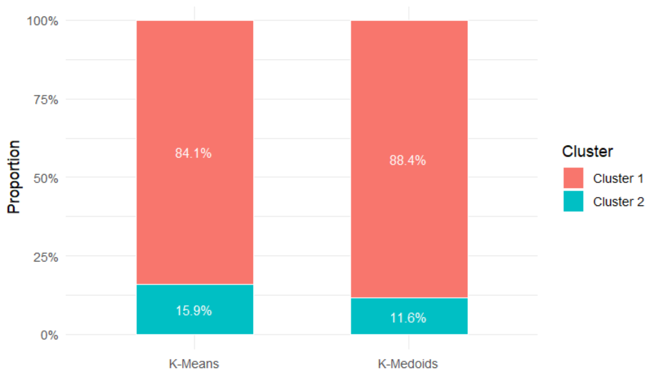

Figure 8. Comparative Composition of Cluster Memberships between K-Means and K-Medoids Methods

Characteristics of Clusters

To examine the differences in characteristics between clusters more thoroughly, a Ridgeline Density analysis was conducted on the distribution of food

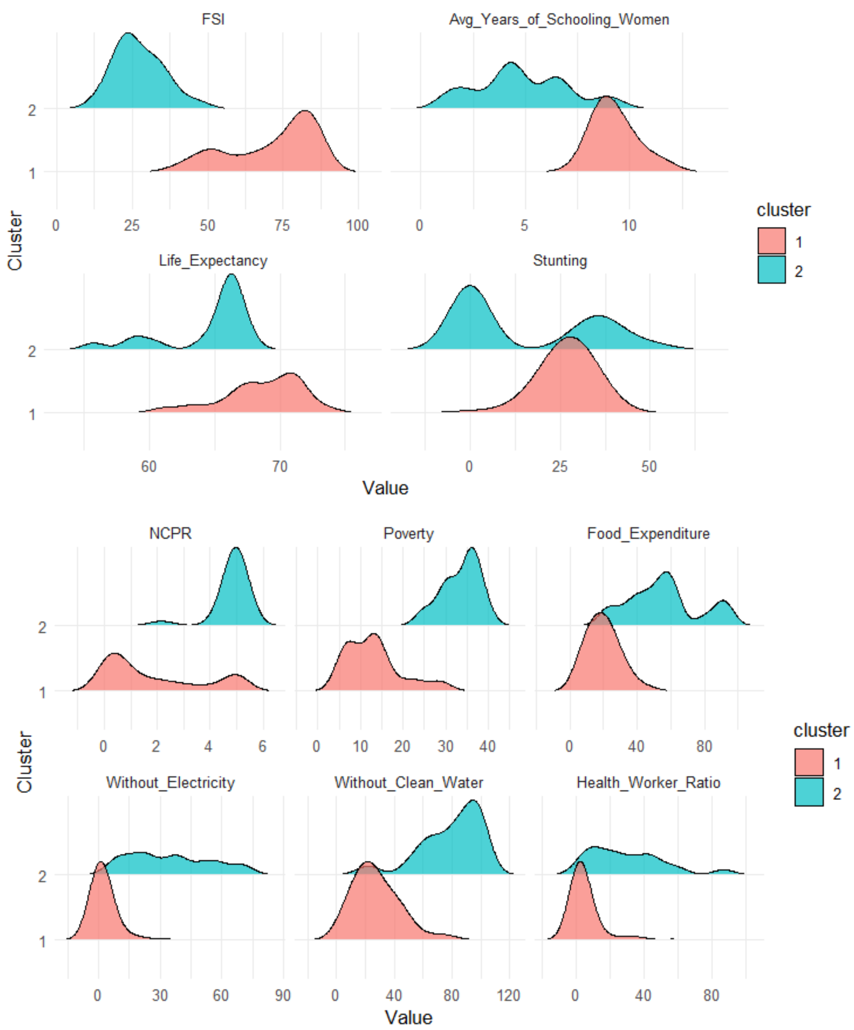

Figure 9. Ridgeline Density Distribution of Food Security Indicators Across Clusters

security indicators. This visualization provides insights into the spread of values, medians, and differences in the mean of key indicators across clusters. The results of this analysis are presented in Figure 9.

Figure 9 illustrates the distribution profile of indicators for each cluster derived from the K-Medoids analysis. It is evident that Cluster 1 (shown in blue) demonstrates relatively better average indicator values compared to Cluster 2. This condition is reflected in higher Food Security Index (FSI) scores, along with supporting life quality indicators such as life expectancy and average years of schooling for women, both of which are at a more favorable level. Moreover, vulnerability-related indicators—such as poverty rate, stunting prevalence, limited access to electricity, and inadequate access to clean water—show lower values in Cluster 1. These characteristics indicate that regions belonging to Cluster 1 generally exhibit stronger and more stable food security conditions.

In contrast, Cluster 2 (in red) represents a more vulnerable condition. The regencies/cities within this cluster are characterized by low FSI, education, and health scores, coupled with higher poverty and stunting rates. Furthermore, the lack of basic infrastructure is more prominent in Cluster 2, as indicated by a higher percentage of households without access to electricity and clean water. These characteristics reinforce that Cluster 2 represents regions with significant food insecurity, particularly in Papua, which consistently faces unfavorable economic, health, and infrastructure conditions.

In conclusion, the differences between Cluster 1 and Cluster 2 are evident not only in overall FSI scores but also across social, economic, educational, and health dimensions. Regions in Cluster 1, which exhibit relatively higher food security and better infrastructure, would benefit from policies focused on strengthening food system sustainability, improving food quality and nutrition, and enhancing regional market integration.

In contrast, Cluster 2, which represents areas with the most severe food insecurity, particularly in Papua, requires targeted interventions addressing basic structural constraints, such as poverty reduction, expansion of electricity and clean water access, improvement of transportation and market connectivity, and strengthening primary healthcare and nutrition services. These cluster-specific policy directions highlight that food security interventions in the Sulampua region should be differentiated based on regional characteristics rather than implemented through a uniform approach.

Conclusion

This study examines food security in the Sulawesi, Maluku, and Papua (Sulampua) regions by integrating multivariate exploration and cluster analysis using the K-Means and K-Medoids methods. The Principal Component Analysis (PCA) reveals that variations in food security are primarily driven by three dimensions: economic vulnerability and access to basic infrastructure, public health and nutrition, and education and demographic conditions. These findings highlight that food security is shaped not only by food availability but also by broader social and human development factors.

The clustering results indicate two distinct regional groups. Cluster 1 consists mainly of regencies and cities in Sulawesi and Maluku, characterized by relatively better yet heterogeneous food security conditions. In contrast, Cluster 2 comprises 16 regencies and cities in Papua, representing regions with the most severe food insecurity. Cluster validation results show that the K-Medoids method outperforms K-Means due to its robustness to outliers and its ability to produce more stable cluster structures.

Overall, this study provides a multivariate and spatially informed perspective on food security disparities in Sulampua. The resulting regional classification offers a practical reference for policymakers to design targeted and cluster-specific food security interventions, particularly for highly vulnerable areas in Papua.

References

Arora, P., Deepali, Varshney, S. 2016. Analysis of K-means and K-medoids algorithm for big data. Procedia Computer Science 78, 507–512.

Aselnino, P., Wijayanto, A.W. 2023. Comparative analysis of hierarchical and non-hierarchical methods in clustering Indonesian provinces based on women empowerment indicators. Indonesian Journal of Applied Statistics, 6(1), 57–68.

Azhar, A.L., Sulianto. 2023. Modeling of food security index in Indonesia based on ordinal logistic regression with random effects panel data approach. Jurnal Ketahanan Nasional 29(2), 166–177. doi:10.22146/jkn.86511.

Badan Pangan Nasional. 2022. Food Security Index. Jakarta: Badan Pangan Nasional.

Dewi, D.A.I.C., Pramita, D.A.K. 2019. Comparative analysis of Elbow and Silhouette methods in K-Medoids clustering algorithm for grouping Balinese handicraft production. Jurnal Matrix, 9(3), 102–108.

Dunn, J.C. 1974. Well separated clusters and optimal fuzzy partitions. Journal of Cybernetics 4(1), 95–104.

Firliana, R., Wulaningrum, R., Sasongko, W. 2015. Implementation of Principal Component Analysis (PCA) for human face recognition. Nusantara of Engineering, 2(1), 65–70.

Food and Agriculture Organization of the United Nations. 2013. The state of food insecurity in the world 2013: The multiple dimensions of food security. Rome: FAO.

Herman, T., Heryanto, R. 2022. Cluster analysis with K-Means versus K-Medoids in financial performance evaluation. Applied Sciences 12(15), 7985.

Khairiyakh, R., Irham, Mulyo, J.H. 2015. Contribution of agricultural sector and sub-sectors to the Indonesian economy. Ilmu Pertanian 18(3), 150–159.

Kossakov, M., Mukasheva, A., Balbayev, G., Seidazimov, S., Mukammejanova, D., Sydybayeva, M. 2024. Quantitative comparison of machine learning clustering methods for tuberculosis data analysis. Engineering Proceedings 60(1), 20. doi:10.3390/engproc2024060020.

Nahdliyah, M.A., Widiharih, T., Prahutama, A. 2019. K-medoids clustering method with silhouette index and C-index validation (Case study: number of crimes in Central Java regencies/cities in 2018). Jurnal Gaussian 8(2), 161–170. Available from: http://ejournal3.undip.ac.id/index.php/gaussian

Prihandoko, Bertalya, Ramadhan, M.I. 2017. An analysis of natural disaster data using K-means and K-medoids algorithms of data mining techniques. Proceedings of the 15th International Conference on Quality in Research (QiR): International Symposium on Electrical and Computer Engineering, 221. IEEE. doi:10.1109/QIR.2017.8168485.

Savira, A.R. 2023. Optimizing clustering models using Principal Component Analysis for car customers. Indonesian Journal of Computing and Cybernetics Systems (IJCCS), 17(2), 145–152. https://doi.org/10.22146/ijccs.94744

Shang, Q., Yu, Y., Xie, T. 2022. A hybrid method for traffic state classification using K-medoids clustering and self-tuning spectral clustering. Sustainability 14(17), 11068. doi:10.3390/su141711068.

Ula, M., Ula, M., Yulisda, D., Susanti, S. 2023. Fuzzy C-means with Borda algorithm in cluster determination system for food-prone areas in North Aceh. ILKOM Jurnal Ilmiah 15(1), 21–31.

Wicaksono, A.P., Widjaja, S., Nugroho, M.F., Christina, Putri, P. 2024. Analysis of K-value toward ticket sales clustering on K-Means using Elbow and Silhouette methods. SISTEMASI: Jurnal Sistem Informasi 13(1), 28–38.

Zhang, L., Yang, Z., Xiu, J., Liu, C. 2020. An optimized interpolation model based on K-means clustering for rainfall calculation. Proceedings of the 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE), 1198–1202. IEEE.