Current Science Research Bulletin
Volume 03 (05), May 2026, pp. 85-92
CrossRef DOI: 10.55677/csrb/01-V03I05Y2026
Available on: https://csrbjournal.org/
e-ISSN: 3050-7103
p-ISSN: 3050-709X

A Comparative Study of Fuzzy C-Means and Kernel K-Means for Clustering Regencies and Cities in Indonesia Based on Food Demand

Guntur Richard Sirait[1], Suparti Suparti[*2], Masithoh Yessi Rochayani[3]
[1,2,3] Department of Statistics, Faculty of Science and Mathematics, Universitas Diponegoro

ABSTRACT: Indonesia's large population leads to high market demand for foods and food products. Because of this, Indonesia faces many challenges in ensuring a sufficient and high-quality food supply. Grouping 514 districts/cities in Indonesia based on food demand aims to effectively understand community needs and adapt marketing strategies to meet market demand. The variables used in this study are the amount of food consumed by districts/cities in Indonesia, which are divided into eleven food groups. The Fuzzy C-Means and Kernel K-Means algorithms were used to group regions based on their food demand. The selection of the optimal method and number of clusters was done by using the Silhouette Coefficient validity index. The optimal cluster is the cluster with the highest Silhouette Coefficient value, closest to 1. Through cluster validity testing using the Silhouette Coefficient, Fuzzy C-Means with three clusters was found to be the most optimal method, having the highest Silhouette Coefficient value, at 0.608436. This indicates that Fuzzy C-Means with three clusters has good cluster distribution. Cluster 1 indicates 88 districts/cities with the highest demand for fruit and other food ingredients. Cluster 2 indicates 224 districts/cities with the highest demand for eggs and milk, vegetables, fish, oil and coconut, beverage ingredients, spices, and processed foods. Cluster 3 indicates 202 districts/cities with the highest demand for meat and nuts.

KEYWORDS: Market Demand; Food Ingredients; Cluster Analysis; Fuzzy C-Means; Kernel K-Means; Silhouette Coefficient

1. INTRODUCTION

Market demand for food is defined as the number of people who want and need food and food products in a given region. Indonesia faces numerous challenges in ensuring a sufficient and high-quality food supply, influenced by Indonesia's status as one of the most populous countries in the world (Azahari, 2008). Regional classification of food demand in Indonesia cannot be ignored as one of the efforts to achieve sustainable food security and effectively meet community needs.

Regional classification of food demand allows producers to understand community preferences, consumption patterns, and identify regional groups and food sources based on market demand. This information is key for planning adequate food production, distribution, and provision (Zakiyah, 2022). Through this understanding, producers can adjust their marketing strategies according to market demand for food.

Clustering is a technique for dividing data into groups so that each group contains similar data or data with similar characteristics. Fuzzy C-Means Clustering and Kernel K-Means Clustering are two frequently used clustering methods. Fuzzy C-Means Clustering groups data into clusters based on the degree of membership of each data item, while Kernel K-Means Clustering is an extension of the K-Means algorithm using kernel methods, allowing high-dimensional data to be mapped into a new feature space and thus linearly separable (Santosa, 2007). The number of clusters to be formed in FCM is determined in advance (Rohmah and Saputro, 2020). Classification doesn't stop there; cluster validation using the silhouette coefficient method is performed to determine cluster strength and determine optimal clusters.

This study uses data on average food consumption per capita by district/city in Indonesia per week in 2022, downloaded from the Central Statistics Agency (BPS) of Indonesia. The data was processed using the FCM and Kernel K-Means Clustering methods, along with cluster validation using the Silhouette Coefficient, and the results were then compared. This study not only clusters food items but also compares the clustering results and selects the best method to generate optimal clusters, resulting in clustering results

Cite the Article: *Sirait, G.R., Suparti, S., Rochayani, M.Y. (2026). A Comparative Study of Fuzzy C-Means and Kernel K-Means for Clustering Regencies and Cities in Indonesia Based on Food Demand. Current Science Research Bulletin, 3(5), 85-92.* https://doi.org/10.55677/csrb/01-V03I05Y2026

Publication Date: *May 14, 2026*

Corresponding Author: Suparti Suparti

that can provide useful information. Interpretation of the clustering results generated through this study is expected to provide insight and knowledge for producers in adjusting their marketing strategies to meet market demand for food items.

2. LITERATURE REVIEW

Clustering is an effort to group data so that each group contains similar data. The goal is to ensure that the data in each cluster can provide useful information. Clustering has become a powerful tool for addressing complex statistical problems (Gustientiedina et al., 2019). When conducting clustering analysis, it is important to ensure that the sample taken is representative of the actual population. This impacts the validity of the analysis results, as they will be generalized to the entire population. Multicollinearity must also be avoided in the data used.

Multicollinearity in data can be detected in various ways, one of which is the Variance Inflation Factor (VIF) value. Multicollinearity in data is indicated by a VIF value ≥ 10, which requires treatment. The VIF value can be obtained using the following formula:

$$VIF_j = \frac{1}{1-R_j^2} \quad (1)$$

where R_j^2 is the coefficient of determination between the independent variables.

The FCM algorithm operates using a fuzzy model, allowing each data point in a group to have varying levels of membership, ranging from 0 to 1 (Bora and Gupta, 2014). The basic concept of FCM begins with determining how many groups you want to form, then determining the cluster center, which indicates the average location for each group. Initially, the position of the cluster center is not precise, so adjustments are made to the cluster center and membership values for each data point through an iterative process. This iterative process is based on an objective function that shows the distance from each data point to the cluster center.

The objective function of the FCM is constrained because it requires a Lagrange multiplier to minimize it. The Lagrange multiplier method can be used to optimize the objective function to find new membership degree parameters and cluster centers (Sanusi et al., 2019).

The objective function of FCM is expressed by the following formula:

$$P_t = \sum_{i=1}^{n} \sum_{k=1}^{c} {d_{ik}}^2 (\mu_{ik})^w \quad (2)$$

with constraint function:

$$\sum_{k=1}^{c} \mu_{ik} = 1 \quad (3)$$

where ${d_{ik}}^2 = \|x_i - v_k\|^2$ is the distance of observation i from the center of cluster k.

The kernel function is defined as a function k for all input vectors $\boldsymbol{x}_i, \boldsymbol{x}_j$ satisfying the conditions:

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \varphi(\boldsymbol{x}_i)^T \varphi(\boldsymbol{x}_j) \quad (4)$$

where φ is the mapping function from the input space to the feature space, or can be written as $\varphi: \boldsymbol{x} \rightarrow \varphi(\boldsymbol{x}) \in F$. Kernel functions can transform a model into a higher-order space without the need to create a function describing the relationship from the input space to the feature space (Maysaroh, 2015).

According to Lόpez et al. (2022), there are four popular basic kernel functions:

a. Linear Kernel

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = {\boldsymbol{x}_i}^T \boldsymbol{x}_j \quad (5)$$

b. Polynomial Kernel

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\gamma {\boldsymbol{x}_i}^T \boldsymbol{x}_j + r)^d, \gamma > 0 \quad (6)$$

c. Gaussian Kernel

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = exp\left\{-\sigma \|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2\right\} \quad (7)$$

d. Exponential kernel

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = tanh(\gamma {\boldsymbol{x}_i}^T \boldsymbol{x}_j + r) \quad (8)$$

where γ, r, and d are kernel indicators, and $i, j = 1,2, \dots, n$.

Corresponding Author: Suparti Suparti

According to Santosa (2007), Kernel K-Means Clustering is an extension of the K-Means algorithm by mapping high-dimensional data into a new space using the kernel method, resulting in more accurate clustering results. The Gaussian kernel is the most commonly used function. The objective function of Kernel K-Means is as follows (Achmal et al., 2022):

$$L_{KKMeans} = \sum_{i=1}^{n} \sum_{k=1}^{c} P_{kn} \, \|\varphi(\boldsymbol{x}_i) - \varphi(\boldsymbol{v}_k)\|^2 \quad (9)$$

$L_{KKMeans}$: objective function of Kernel K-Means
P_{kn} : membership value
$\|\varphi(\boldsymbol{x}_i) - \varphi(\boldsymbol{v}_k)\|^2$: new distance of the i-th observation from the k-th cluster center in kernel space

The silhouette coefficient is a cluster validation technique using internal criteria to measure the extent to which an object is within its intended cluster. Cluster validation using internal criteria examines the distance between clusters and between clusters. Optimal clusters are those with small distance differences within clusters and large distance differences with other clusters. The silhouette coefficient depends only on the placement of objects within each cluster, not on the clustering algorithm used.

The following are the steps in calculating the silhouette coefficient (Struyf et al., 1997):

1. Calculate the average distance between the i-th observation and all observations in a cluster.

$$a(i) = \frac{1}{|A| - 1} \sum_{j \in A, j \neq 1} d(i,j) \quad (10)$$

$a(i)$: the average distance between the i-th observation and all observations in cluster A
$d(i,j)$: the distance between observations i and j
$|A|$: the number of observations in cluster A

2. Calculate the average distance between the i-th object's observations and all observations in other clusters.

$$d(i,C) = \frac{1}{|C|} \sum_{j \in C} d(i,j) \quad (11)$$

$d(i,C)$: average distance of observation i with all observations in cluster C with C ≠ A
$|C|$: number of observations in cluster C

3. Calculate $d(i, C)$ for all C and then determine the minimum value.

$$b(i) = min\, d(i,C) \quad (12)$$

$b(i)$: the minimum value of the average distance between observation i and all observations in cluster C

4. Calculate the silhouette coefficient value

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}} \quad (13)$$

$s(i)$: silhouette coefficient value

The interpretation of the silhouette coefficient value is shown in Table 1.

Table 1. Interpretation of Silhouette Coefficient Values

Silhouette Coefficient Values	Interpretation
0.70≤ SC ≤1.00	Strong cluster distribution
0.50≤ SC <0.70	Good/standard cluster distribution
0.20≤ SC <0.50	Weak cluster distribution
SC ≤0.20	Unstructured cluster distribution

3. DATA AND METHODOLOGY

The data for this study were obtained from the 2023 publication of the Indonesian Central Statistics Agency (BPS) on Average Food Consumption Per Capita in One Week by Regency/City in Indonesia in 2022, covering 514 regencies/cities. The variables used were 11 food groups consumed by regencies/cities in Indonesia, as shown in Table 2. This study used R Studio and Microsoft Excel for the analysis process, with the following steps:

1. Preparing the data for analysis.
2. Conducting cluster analysis assumption tests using representative samples and multicollinearity tests using VIF values.
3. Data analysis using the FCM algorithm

Corresponding Author: Suparti Suparti

a. Determine the initial parameters, namely the number of *cluster* (c), weighting (w), maximum iteration (*MaxIter*), desired error (ε), initial objective function ($P_0 = 0$), and initial iteration ($t = 1$)

b. Generate random numbers for the members of the initial partition matrix U

$$\boldsymbol{U} = \begin{bmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1c} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2c} \\ \vdots & \vdots & \vdots & \vdots \\ \mu_{n1} & \mu_{n2} & \cdots & \mu_{nc} \end{bmatrix} \quad (14)$$

with the condition:

$$\sum_{k=1}^{c} \mu_{ik} = 1$$

c. Calculating the cluster center of the t-th iteration

$$V_{kj} = \frac{\sum_{i=1}^{n} (\mu_{ik})^w x_{ij}}{\sum_{i=1}^{n} (\mu_{ik})^w} \quad (15)$$

d. Calculating the objective function of the t-th iteration

$$P_t = \sum_{i=1}^{n} \sum_{k=1}^{c} \left(\left[\sum_{j=1}^{p} (x_{ij} - v_{kj})^w \right] (\mu_{ik})^w \right) \quad (16)$$

e. Calculating the change in membership degree of each cluster

$$\mu_{ik} = \frac{\left[\sum_{j=1}^{p} (x_{ij} - v_{kj})^2 \right]^{\frac{-1}{w-1}}}{\sum_{k=1}^{c} \left[\sum_{j=1}^{p} (x_{ij} - v_{kj})^2 \right]^{\frac{-1}{w-1}}} \quad (17)$$

f. Checking the stopping condition
   - ($|P_t - P_{t-1}| < \varepsilon$) or ($t >$ maximum iteration) then stop.
   - ($|P_t - P_{t-1}| > \varepsilon$) or ($t <$ maximum iteration) then $t = t + 1$, repeat the cluster center calculation.

g. Validate the cluster using the silhouette coefficient to determine the optimal cluster.

4. Group districts/cities based on food demand based on the FCM algorithm.
5. Data analysis using the Kernel K-Means Clustering algorithm
   a. Form an N×N kernel matrix K
   b. Determine the eigenvalues of the kernel matrix and then determine the eigen vectors corresponding to the eigen values
   c. Normalize all H matrices to the unit sphere
   d. Determine the Kernel K-Means clustering of the matrix
   e. Validate clusters using the Silhouette Coefficient to determine optimal clusters
6. Group districts/cities based on food demand based on the Kernel K-Means Clustering algorithm.
7. Interpret and compare the clustering results between FCM and Kernel K-Means Clustering.

Table 2. Research Variables

Variables	Informations	Units
X1	Meat group	kg
X2	Egg and Milk group	kg
X3	Vegetable group	kg
X4	Fish Group	kg
X5	Nuts Group	kg
X6	Fruit group	kg
X7	Oil and Coconut group	kg
X8	Beverage ingredients group	kg
X9	Other food ingredients group	kg
X10	Spices group	kg
X11	Prepared food and beverages group	kg

Corresponding Author: Suparti Suparti

4. RESULTS AND DISCUSSION

The research began with an assumption test on the data to be used. In cluster analysis, data must meet the assumptions of sample representativeness and the assumption of multicollinearity. This study did not conduct a sample representativeness test because the research data already comprised population data from all districts/cities in Indonesia. The multicollinearity assumption test was conducted using the Variance Inflation Factor (VIF). The assumption is considered satisfied if the VIF value of each tested variable is less than 10. Based on the results from R Studio software, the VIF values for each variable are as shown in Table 3.

Table 3. Multicollinearity Assumption Test with VIF

Variables	Informations	VIF Values
X1	Meat group	2.207943
X2	Egg and Milk group	4.587065
X3	Vegetable group	1.720248
X4	Fish Group	1.373030
X5	Nuts Group	2.515210
X6	Fruit group	1.274695
X7	Oil and Coconut group	1.413986
X8	Beverage ingredients group	1.243022
X9	Other food ingredients group	2.495921
X10	Spices group	1.344486
X11	Prepared food and beverages group	2.041373

Based on Table 3, the VIF values for all variables are less than 10, thus concluding that the correlation is moderate, thus meeting the multicollinearity assumption, allowing the cluster analysis to proceed.

The FCM analysis begins by determining the initial parameter values. In this study, the clustering process was carried out with three clusters. The initial parameters used were as follows:

1. Number of clusters (c) : 3
2. Weight (w) : 2
3. Maximum iterations : 1000
4. Minimum error ($\mathcal{E}$) : 10^{-4}
5. Initial objective function (P_0) : 0
6. Initial iteration (t) : 1

Next, generate random numbers to form the U matrix as the initial membership degree, provided that each row must have a value of 1 ($\sum_{k=1}^{c} \mu_{ik} = 1$). The random number generation was achieved using R Studio software, as shown in Table 4.

Table 4. Random Number Generation Results

No	*Cluster 1*	*Cluster 2*	*Cluster 3*
1	0.844295	0.091286	0.064419
2	0.992113	0.002503	0.005384
3	0.30882	0.620209	0.070971
⋮	⋮	⋮	⋮
513	0.676978	0.191139	0.131883
514	0.038704	0.006013	0.955283

The initial cluster center calculation is based on the number of clusters to be formed. The cluster center values for the first iteration are as shown in Table 5.

Table 5. Cluster Center Calculation Results 1-3 First Iterations

v_{1j}	0.197	2.019	1.811	0.974	0.108	0.546	1.032	3.497	1.002	78.54	7.035
v_{2j}	0.138	2.037	8.365	4.44	0.083	0.222	8.66	32.456	0.749	92.423	9.353
v_{3j}	0.258	1.881	0.54	0.133	0.137	0.229	0.721	5.534	0.172	9.772	0.033

The objective function is calculated to obtain accurate cluster centers. Iterations are performed repeatedly until the objects are placed in the appropriate clusters. Before calculating the objective function, the distance between the data objects (X_{ij}) and the cluster centers (V_{kj}) is calculated. The objective function for the first iteration is as shown in Table 6.

Table 6. Calculation of the Objective Function of the First Iteration

i	${\mu_{i1}}^2 \times \sum_{j=1}^{11}(x_{ij}-v_{1j})^2$	${\mu_{i2}}^2 \times \sum_{j=1}^{11}(x_{ij}-v_{2j})^2$	${\mu_{i3}}^2 \times \sum_{j=1}^{11}(x_{ij}-v_{3j})^2$	P_t
1	22.33295	8.358341	22.29761	52.9889
2	21.52763	0.007927	0.125928	21.66149
3	21.9901	396.8652	35.78595	454.6413
⋮	⋮	⋮	⋮	⋮
514	0.910159	0.087668	1827.119	1828.116
	$P_t = \sum_{i=1}^{514}\sum_{k=1}^{3}\left(\left[\sum_{j=1}^{p}(X_{ij}-V_{kj})^2\right](\mu_{ik})^2\right)$			235745.8

Thus, the objective function value obtained in the first iteration was 235745.8.

Next, the change in the membership degree value for each data item in all clusters was calculated. The new membership degree results for the first iteration are shown in Table 7.

Table 7. Calculation of the New Degree of Membership of the First Iteration

i	μ_{i1}	μ_{i2}	μ_{i3}
1	0.964259	0.030119	0.005622
2	0.978173	0.016902	0.004925
3	0.796216	0.177943	0.025841
⋮	⋮	⋮	⋮
514	0.64348	0.161247	0.195273

The stopping condition is checked by comparing the value of P_t with P_{t-1}. Based on the previous calculation, the objective function value is 235745.8. The ratio of $|P_t - P_{t-1}|$ to ε is $|P_1 - P_0| = |235745.8 - 0| = 235745.8 > 10^{-4}$. Since the value of $|P_t - P_{t-1}| > \varepsilon$, the iteration continues with $t = t + 1$. Based on data analysis using R Studio software, the iteration process for Fuzzy CMeans Clustering with 3 clusters was performed 61 times.

Kernel K-Means Clustering analysis begins with deciding on the number of clusters to be formed, namely 3 clusters, and the free parameter (σ) value of 0.01. A feature space matrix of size ($N \times N$) must be formed before the analysis. The elements of the feature space matrix are calculated using a Gaussian kernel. The following is the resulting kernel matrix.

$$K = \begin{bmatrix} 1 & 0.821625 & 0.928465 & \cdots & 0.602941 & 0.760456 \\ 0.821625 & 1 & 0.890886 & \cdots & 0.685931 & 0.848836 \\ 0.928465 & 0.890886 & 1 & \cdots & 0.625423 & 0.747274 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0.602941 & 0.685931 & 0.625423 & \cdots & 1 & 0.772035 \\ 0.760456 & 0.848836 & 0.747274 & \cdots & 0.772035 & 1 \end{bmatrix}$$

Next, calculate the eigenvalues and eigenvectors of the kernel matrix with three clusters using the R Studio syntax. The results are as follows:

Eigen values: 418.7039; 25.49934; 13.3029; 11.16942;⋯; 3.6451×10^{-7} ; 3.1646×10^{-7}

Eigen vectors:

$$\begin{bmatrix} -0.041151 & -0.055345 & -7.529229 \\ -0.045283 & -0.003539 & -5.121725 \\ -0.041076 & -0.049204 & -6.717407 \\ \vdots & \vdots & \vdots \\ -0.036642 & -0.009101 & -0.023176 \\ -0.046800 & 0.000345 & 0.052223 \end{bmatrix}$$

Normalizing all H matrices to the unit sphere is done by running the syntax in R Studio software as follows:

$$\begin{bmatrix} -0.403028 & -0.542044 & -0.737398 \\ -0.661490 & 0.051702 & -5.748169 \\ -0.442406 & -0.529944 & -0.723487 \\ \vdots & \vdots & \vdots \\ -0.827109 & -0.205445 & -0.523145 \\ -0.667373 & 0.004921 & 0.744706 \end{bmatrix}$$

The formed clusters were then validated using the silhouette coefficient to determine the best cluster. The silhouette coefficient calculation was performed using R Studio software, yielding the validation results as shown in Table 8.

Table 8. Cluster Validity Index

Number of Clusters	Silhouette Coefficient Index	
	Fuzzy C-Means	Kernel K-Means
2	0.583271	0.139416
3	0.608436	0.154802
4	0.596182	0.109729
5	0.550593	0.078379
6	0.560605	0.046071
7	0.525839	0.038128
8	0.490288	0.013169

Table 8 shows that the FCM method is superior because it has a higher silhouette coefficient value than the Kernel K-Means method across all clusters. The highest silhouette coefficient value was 0.608436 for FCM with 3 clusters, so the best cluster is the one formed using the FCM method with 3 clusters.

The cluster profiling stage involves naming or labeling the clusters to identify the characteristics of each cluster. Profiling can be performed by measuring the centroid values. The cluster profiling results are presented in Table 9.

Table 9. Centroid Values of Variables in Each Cluster

Cluster	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	Number of members
1	0.197	2.019	1.811	0.974	0.108	0.546	1.032	3.497	1.002	78.540	7.035	88
2	0.138	2.037	8.365	4.440	0.083	0.222	8.660	32.456	0.749	92.423	9.353	224
3	0.258	1.881	0.540	0.133	0.137	0.229	0.721	5.534	0.172	9.772	0.033	202

Based on Table 9, the following are the interpretations of the average values (characteristics) that appear in each cluster.

1. *Cluster* 1

Cluster 1 obtained higher average values for the fruit group (X6) and other food ingredients (X9) compared to other clusters. Cluster 1 consisting of 88 districts/cities, has the highest demand for food ingredients in the fruit group and other food ingredients.

2. *Cluster* 2

Cluster 2 obtained higher average values for the egg and milk group (X2), vegetable group (X3), fish group (X4), oil and coconut group (X7), beverage group (X8), spices group (X10), and processed food and beverage group (X11) compared to other clusters. Cluster 2 consisting of 224 districts/cities, has the highest demand for food ingredients in the egg and milk group, vegetable group, fish group, oil and coconut group, beverage group, spices group, and processed food and beverage group.

3. *Cluster* 3

Cluster 3 obtained higher average scores for the meat (X1) and legume (X5) variables compared to the other clusters. Cluster 3 comprising 202 districts/cities, has the highest demand for meat and legume foods. Based on the cluster interpretation results above, the parties involved can consider various related issues as an effort to achieve sustainable food security and effectively meet community needs and market demand for food.

5. CONCLUSIONS

Through the analysis and discussion of the two applied methods, it can be concluded that cluster analysis using the Fuzzy C-Means (FCM) and Kernel K-Means algorithms on the average per capita weekly food consumption data of regencies/cities in Indonesia in 2022, evaluated using the silhouette coefficient, produced validity index values of 0.608436 for FCM and 0.154802 for Kernel K-Means. These results indicate that the FCM method performs better than Kernel K-Means, as it yields a higher silhouette coefficient across all clusters. Furthermore, the optimal number of clusters identified using the FCM method is three, as it provides a higher silhouette coefficient compared to alternative cluster configurations.

Based on the profiling results, Cluster 1, which consists of 88 regencies/cities, shows the highest demand for food items in the fruit and other food groups. Cluster 2, comprising 224 regencies/cities, demonstrates the highest demand for eggs and milk, vegetables, fish, oils and coconut-based products, beverage ingredients, spices, as well as processed food and beverages. Meanwhile, Cluster 3, consisting of 202 regencies/cities, exhibits the highest demand for food items in the meat and legumes groups.

REFERENCES

1. Achmal, E. F., Cholissodin, I., & Adikara, P. P. (2022). Segmentasi pelanggan menggunakan metode kernel K-means (studi kasus: Smartlegal.id). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 6(6), 2835–2843.
2. Azahari, D. H. (2008). Membangun kemandirian pangan dalam rangka meningkatkan ketahanan nasional. *Analisis Kebijakan Pertanian*, 6(2), 174–195.
3. Bora, D. J., & Gupta, D. A. K. (2014). A comparative study between fuzzy clustering algorithm and hard clustering algorithm. *International Journal of Computer Trends and Technology*, 10(2), 108–113
4. Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan algoritma K-means untuk clustering data obat-obatan. *Jurnal Nasional Teknologi dan Sistem Informasi*, 5(1), 17–24.
5. López, O. A. M., López, A. M., & Crossa, J. (2022). *Multivariate statistical machine learning methods for genomic prediction*. Springer Nature.
6. Maysaroh, S. (2015). *Analisis kelompok dengan metode kernel K-means (studi kasus pengelompokan kabupaten/kota di Indonesia berdasarkan penduduk dengan faktor-faktor risiko penyebab penyakit hipertensi)* (Tesis). Institut Teknologi Sepuluh Nopember.
7. Rohmah, D. S., & Saputro, D. R. S. (2020). Clustering data dengan algoritme fuzzy C-means berbasis indeks validitas partition coefficient and exponential separation (PCAES). In *PRISMA, Prosiding Seminar Nasional Matematika* (Vol. 3, pp. 58–63).
8. Santosa, B. (2007). *Data mining terapan dengan Matlab*. Graha Ilmu.
9. Sanusi, W., Zaki, A., & Afni, B. N. (2019). Analisis fuzzy C-means dan penerapannya dalam pengelompokan kabupaten/kota di Provinsi Sulawesi Selatan berdasarkan faktor-faktor penyebab gizi buruk. *Journal of Mathematics, Computations and Statistics*, 2(1), 47–54.
10. Struyf, A., Hubert, M., & Rousseuw, P. J. (1997). Integrating robust clustering techniques in S-PLUS. *Computational Statistics & Data Analysis*, 26(1), 17–37.
11. Zakiah, S. (2022). Teori konsumsi dalam perspektif ekonomi Islam. *El-Ecosy: Jurnal Ekonomi dan Keuangan Islam*, 2(2), 180–194.

Corresponding Author: Suparti Suparti