International Journal of Advances in Data and Information Systems
Vol. 7, No. 3, December 2026, pp. 1030~1053
ISSN: 2721-3056, DOI: 10.59395/ijadis.v7i3.1661

A Comparative Study of Clustering Algorithms for Indonesian Provincial Food Security

Ika Nur Laily Fitriana[1], Emeylia Safitri[2], Siti Hadijah Hasanah[3]
[1,2,3]Statistics Study Program, Universitas Terbuka, Indonesia

Article Info

Article history:

Received Jun 30, 2026
Revised Aug 18, 2026
Accepted Sept 29, 2026

Keywords:

Food Security
Clustering Algorithms
Internal Validity Indices
Bootstrap Stability Analysis
Provincial Typology

ABSTRACT

Food security disparities in Indonesia remain multidimensional and spatially uneven, requiring analytical approaches that can capture regional differences beyond single indicators. This study compares multiple clustering algorithms to classify 38 Indonesian provinces using twelve food security indicators representing food availability, affordability, utilization, and consumption outcomes. Provincial secondary data for 2024 were obtained from official sources, including Statistics Indonesia, the National Food Agency, and the Ministry of Health. The analysis began with data standardization and Principal Component Analysis, which retained three components explaining 74.65% of total variance. Ten clustering approaches were then evaluated, including K-Means, hierarchical linkage methods, Fuzzy C-Means, DBSCAN, HDBSCAN, Gaussian Mixture Model, and ensemble clustering. Each solution was assessed using internal validity indices, including Silhouette, Davies–Bouldin, and Dunn Index, as well as degeneracy checks, cluster membership consistency, bootstrap stability analysis based on ARI and NMI, robustness checks under alternative preprocessing and variable subsets, and substantive interpretability. Ward Linkage was selected as the final method because it produced a non-degenerate and interpretable three-cluster typology, supported by similar partitions from Complete Linkage and ensemble clustering. The two-province high-vulnerability cluster persisted under log-transformation of skewed indicators, removal of individual variables, and reduced component dimensionality. The final clusters reveal clear territorial disparities: a relatively food-secure cluster covering most western and central provinces, a moderately vulnerable eastern cluster, and a high-vulnerability cluster consisting of Papua Pegunungan and Papua Selatan. Poverty, electricity access, protein consumption, health service burden, stunting, and life expectancy were the strongest differentiating indicators. These findings demonstrate the value of combining internal validity and stability analysis to support robust provincial food security typologies and targeted regional policy interventions.

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

Corresponding Author:

Ika Nur Laily Fitriana,
Statistics Study Program,
Universitas Terbuka,
Tangerang Selatan, Banten, 15437
Email: ika.nur@ecampus.ut.ac.id

1. INTRODUCTION

Food security remains a multidimensional development challenge that cannot be adequately represented by production or supply indicators alone. The Food and Agriculture Organization (FAO) of the United Nations has consistently emphasized that food security comprises four interlinked dimensions, namely availability, access, utilization, and stability, and that progress at the aggregate level may conceal substantial subnational disparities [1]. Barrett [2] argues that food insecurity measurement should account for the multiple pathways through which households and regions experience constraints in availability, access, and utilization, rather than relying on a single proxy. Jones et al. [3] show that different food security metrics capture different facets of the problem and that policy conclusions may vary depending on the indicators and analytical approach adopted. This perspective is further strengthened by [4], who conceptualize food security as a dynamic and resilience-related condition rather than a static outcome, implying that any classification of food-secure versus food-insecure populations must remain sensitive to changing socioeconomic and environmental circumstances.

In Indonesia, food security has become an important policy issue, and its multidimensional framing has direct policy relevance. Although national-level indicators suggest progress in several aspects of food security, official statistics and food security assessments continue to show substantial disparities across provinces, particularly between western and eastern Indonesia. Provinces in eastern Indonesia are more likely to be vulnerable across dimensions related to food availability, affordability, and utilization, indicating that food security problems are not distributed randomly across space but are shaped by structural and territorial inequalities [5]. This makes provincial-level analysis important.

Previous studies have shown that food insecurity in Indonesia is strongly associated with socioeconomic characteristics, household welfare, and access to resources. Amrullah et al. [6] provide important evidence on the household groups most exposed to food insecurity in Indonesia, while [7] shows that socioeconomic characteristics significantly shaped household food security during the COVID-19 pandemic. Further evidence shows that the pandemic intensified food insecurity among Indonesian households, highlighting the sensitivity of food security to economic and social shocks [8]. These studies make important contributions by identifying determinants of food insecurity, yet they primarily focus on household-level explanations and do not fully address how provinces can be grouped into interpretable regional typologies based on multiple food security indicators.

Several studies have attempted to map food security conditions using spatial or clustering approaches. A study by [9] identifies geographic clusters and regional disparities in Indonesian food security using spatial analysis, providing evidence that food insecurity is spatially patterned rather than randomly distributed. A model-based clustering approach was applied in [10], showing that provincial food security conditions can be represented through a three-cluster structure. This direction is further developed in [11] through a cluster ensemble-based mixed-data approach, demonstrating that clustering can support food security mapping when indicators have heterogeneous data types and scales. These studies establish that unsupervised learning is a viable strategy for provincial food security typology. Nevertheless, most existing studies rely on a single clustering method or a limited family of methods. As a result, they do not fully address whether alternative algorithms would generate comparable partitions and rarely evaluate the resulting partitions using a comprehensive comparison of internal validity, degeneracy, membership consistency, and bootstrap stability. Prior work [12] cautions that cluster validity indices should not be used mechanically, as different indices may favor different structures. Another study [13] further emphasizes that cluster stability analysis is necessary to assess whether clustering results remain consistent under data perturbation or resampling. Therefore, selecting a clustering method for policy-oriented regional typology requires more than identifying the highest internal validity score; it also requires evaluating whether the partition is non-degenerate, stable, interpretable, and substantively meaningful.

The novelty of this study lies in three contributions. First, it develops a layered selection framework in which internal validity indices are treated as necessary but not sufficient evidence: candidate partitions are additionally screened for degeneracy, cross-method membership

consistency, bootstrap stability, robustness to preprocessing and variable selection, and substantive interpretability. This framework is applied to ten clustering algorithms spanning partitional, hierarchical, fuzzy, mixture-model, density-based, and ensemble families, with Principal Component Analysis used beforehand to reduce the dimensionality of the twelve food security indicators. Second, the study shows empirically why such layering is necessary: in this dataset the algorithms attaining the highest raw validity scores produce degenerate partitions that separate only one or two provinces while leaving the remainder undifferentiated and would therefore have been selected by an evaluation based on validity indices alone. Third, the analysis provides provincial evidence for Indonesia's current 38-province configuration, including the four provinces established in 2022 through the division of Papua, which did not exist as separate units when earlier provincial food security studies were conducted. Because the most vulnerable cluster identified here consists of two of these newly established provinces, the results offer an initial baseline for areas that were previously aggregated within a single provincial unit. This study adapts the FAO food security framework by operationalizing availability, affordability (as a proxy for access), utilization, and consumption outcomes. The stability dimension is not directly measured because comparable provincial-level indicators were not available in the compiled dataset.

Accordingly, this study aims to classify Indonesian provinces into meaningful food security typologies by comparing alternative clustering algorithms and identifying the most appropriate method for provincial profiling. Specifically, this study seeks to: (1) compare the performance and stability of multiple clustering algorithms; (2) select the most interpretable and non-degenerate clustering solution; and (3) characterize provincial clusters based on food availability, affordability, utilization, and consumption indicators. The findings are expected to provide a data-driven basis for identifying regional disparities and supporting more targeted food security policy interventions in Indonesia.

2. RESEARCH METHOD

2.1 Data and Research Variables

This study used provincial-level secondary data for the year 2024, covering 38 provinces in Indonesia. The data were obtained from official sources, including Statistics Indonesia (BPS), the National Food Agency (BAPANAS), and the Ministry of Health. The research variables are shown in Table 1.

Table 1. Research Variables

Variables	Dimension	Description	Unit
X1	Food availability	Ratio of normative per-capita consumption to food production	ratio
X2	Food affordability	Population living below the poverty line	%
X3	Food affordability	Households with food expenditure exceeding 65% of total expenditure	%
X4	Food affordability	Households without access to electricity	%
X5	Food utilization	Average years of schooling of females aged 15 years and above	years
X6	Food utilization	Households without access to clean water	%
X7	Food utilization	Population-to-health-worker ratio adjusted for population density	ratio
X8	Food utilization	Prevalence of stunting among children under five years of age	%
X9	Food utilization	Life expectancy at birth	years
X10	Food consumption	Energy consumption per capita per day	kcal/cap/day
X11	Food consumption	Protein consumption per capita per day	g/cap/day
X12	Food availability	Rice productivity	ton/ha

2.2 Research Workflow

This study was conducted through the following analytical steps:

1. Data collection: provincial-level food security indicators are collected and structured from the websites of official sources as the main dataset.
2. Data exploration: descriptive statistics and boxplots are conducted to examine data patterns.
3. Data standardization: all indicators are standardized using Z-scores to ensure comparability across different units and scales.

4. Principal Component Analysis: PCA is applied to reduce dimensionality, and the retained component scores are used for clustering.
5. Determining optimal K: candidate numbers of clusters are evaluated using the elbow method based on within-cluster sum of squares (WCSS), complemented by the average Silhouette width, the Calinski–Harabasz index, and the gap statistic.
6. Multi-method clustering: ten clustering algorithms are applied to compare alternative provincial food security typologies, namely K-Means, Single Linkage, Complete Linkage, Average Linkage, Ward Linkage, Fuzzy C-Means, DBSCAN, HDBSCAN, Gaussian Mixture Model (GMM), and Ensemble Clustering.
7. Cluster validity: clustering results are evaluated using internal validity indices, such as Silhouette, Davies–Bouldin, and Dunn Index.
8. Cluster membership by method: cluster assignments from each method are compared to assess membership consistency.
9. Clustering stability analysis: bootstrap-based stability analysis is conducted using the Adjusted Rand Index (ARI) and Normalized Mutual Information (NMI).
10. Robustness analysis: the sensitivity of the final typology to preprocessing choices, component dimensionality, and variable subsets is examined.
11. Final method selection: the best clustering method is selected based on validity, stability, consistency, and interpretability.
12. Cluster profiling and interpretation: final clusters are profiled using the original indicators to interpret provincial food security typologies.

A comprehensive overview of the research methodology is presented in Figure 1, which illustrates the analytical workflow of this study.

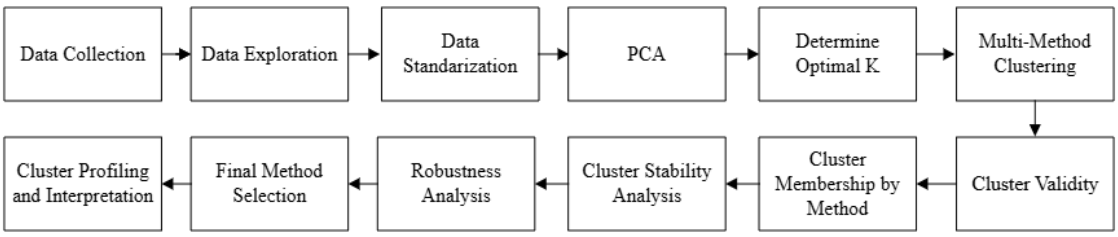


Figure 1. Flow Chart

2.3 Principal Component Analysis (PCA)

PCA is widely used in multivariate analysis to transform correlated variables into uncorrelated components that summarize the main sources of variance [14]. In clustering analysis, PCA is commonly used as a preprocessing step before applying clustering algorithms. By projecting the original variables into a lower-dimensional component space, PCA can reduce noise, simplify the distance structure, and improve the interpretability of clustering results. The component scores obtained from PCA can then be used as input for clustering, allowing the grouping process to be based on the dominant patterns of variation rather than on the full set of correlated original variables [15]. The number of retained components was assessed using the Kaiser eigenvalue criterion, the cumulative proportion of explained variance, and Horn's parallel analysis, which compares each observed eigenvalue with the distribution of eigenvalues obtained from random data of the same dimension [16].

2.4 Clustering Algorithms

Clustering is an unsupervised learning approach used to group observations based on similarity or dissimilarity without predefined class labels. Since no clustering method is universally optimal for all data structures, comparison across algorithm families is recommended in exploratory unsupervised learning [17], [18]. In this study, ten clustering algorithms were

considered, covering partitional, hierarchical, fuzzy, density-based, model-based, and ensemble approaches. The ten clustering algorithms used are as follows:

1. K-Means

K-Means is a partitional clustering algorithm that divides observations into a predetermined number of clusters by minimizing within-cluster variation. This method is widely used because of its computational efficiency and interpretability, although it is more suitable for compact and relatively spherical clusters and can be sensitive to initial centroid selection [18], [19].

2. Single Linkage

Single Linkage is a hierarchical agglomerative clustering method that defines the distance between two clusters using the shortest distance between their members. This method can identify elongated cluster structures, but it may be affected by chaining effects when observations are connected through close intermediate points [18].

3. Complete Linkage

Complete Linkage is a hierarchical clustering method that defines the distance between two clusters using the farthest distance between their members. Compared with Single Linkage, this method tends to produce more compact clusters because cluster merging depends on the maximum inter-object distance [18].

4. Average Linkage

Average Linkage is a hierarchical clustering method that calculates the distance between two clusters based on the average pairwise distance between all observations in the two clusters. This method provides a compromise between Single and Complete Linkage by considering the overall distance structure rather than only the nearest or farthest observations [18].

5. Ward Linkage

Ward Linkage is a hierarchical clustering method that merges clusters by minimizing the increase in within-cluster variance. This method is often used for quantitative data because it tends to produce compact and internally homogeneous clusters [18].

6. Fuzzy C-Means

Fuzzy C-Means is a soft clustering algorithm that allows each observation to have partial membership in more than one cluster. This approach is useful when cluster boundaries are not clearly separated, since each observation is assigned a degree of membership rather than a single exclusive label [18].

7. Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

DBSCAN is a density-based clustering algorithm that identifies clusters as dense regions separated by lower-density areas. It can detect non-spherical clusters and identify noise or outlying observations, making it useful when the data structure does not follow centroid-based assumptions [18].

8. Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN)

HDBSCAN extends the density-based clustering principle by constructing a hierarchical structure of density-connected observations. It is more flexible than DBSCAN when clusters have varying densities and when the number of clusters is not specified in advance [18].

9. Gaussian Mixture Model (GMM)

Gaussian Mixture Model is a model-based clustering approach that assumes observations are generated from a mixture of probability distributions. Each cluster is represented by a Gaussian component, and membership can be interpreted probabilistically, which makes this method useful when cluster uncertainty needs to be considered [18].

10. Ensemble Clustering

Ensemble Clustering combines multiple clustering results to obtain a consensus partition. This method is useful when different algorithms produce different cluster assignments, because it seeks a more robust clustering structure by integrating information from multiple partitions [17].

In this study, the ensemble was implemented as evidence-accumulation (co-association) clustering within the cluster-ensemble framework of [20]. The base ensemble comprised the seven

non-density-based partitions produced by K-Means, the four hierarchical linkages, Fuzzy C-Means, and GMM; DBSCAN and HDBSCAN were excluded because their noise assignments would leave the co-association evidence incomplete. The exact procedure is given in Algorithm 1.

Algorithm 1. Consensus clustering via the co-association matrix

Input: the retained principal component scores of the n = 38 provinces; the M = 7 base algorithms (K-Means, Single, Complete, Average, and Ward Linkage, Fuzzy C-Means, GMM); the number of clusters k = 3.

Step 1. Fit each base algorithm to the component scores to obtain M base partitions.

Step 2. Form the n × n co-association matrix S, whose entry S(i, j) is the fraction of the M base partitions that assign provinces i and j to the same cluster.

Step 3. Convert S into a dissimilarity matrix D = 1 − S.

Step 4. Apply Ward agglomerative clustering to D and cut the resulting tree into k = 3 clusters.

Output: the consensus partition of the 38 provinces.

2.5 Cluster Validity and Stability Evaluation

Cluster validity was evaluated using the Silhouette coefficient, Davies–Bouldin Index, and Dunn Index. The Silhouette coefficient measures how well an observation fits within its assigned cluster relative to the nearest alternative cluster [21]

$$s(i) = \frac{\big(b(i) - a(i)\big)}{\max\{a(i), b(i)\}} \tag{1}$$

where $a(i)$ represents the average distance between observation (i) and all other points within the same cluster, while $b(i)$ denotes the minimum average distance between observation (i) and points in the nearest neighboring cluster. A higher Silhouette value indicates better cluster assignment.

The Davies–Bouldin Index evaluates the ratio between within-cluster scatter and between-cluster separation, where lower values indicate better clustering performance [22]:

$$DBI = \frac{1}{k}\sum_{c=1}^{k} \max_{d \neq c}\left(\frac{S_c + S_d}{M_{cd}}\right) \tag{2}$$

where k is the number of clusters, S_c and S_d represent the average within-cluster distances of clusters c and d, respectively, and M_{cd} denotes the distance between the centroids of clusters c and d. A smaller DBI value indicates more compact clusters that are well separated from each other.

The Dunn Index measures the ratio between minimum inter-cluster separation and maximum intra-cluster diameter, where higher values indicate better-defined clusters [23]

$$Dunn = \frac{min_{1 \le c < d \le k}\delta(C_c, C_d)}{max_{1 \le c \le k}\Delta(C_c)} \tag{3}$$

where $\delta(C_c, C_d)$ represents the distance between clusters C_c and C_d, and $\Delta(C_c)$ denotes the intra-cluster compactness of cluster C_c. Higher Dunn Index values indicate better-defined clusters.

Because internal validity indices may favor different structures and should not be used mechanically [12], the final method was selected through an explicit hierarchical decision rule applied in a fixed order. Rule 1 (non-degeneracy): solutions exhibiting chaining, single-cluster collapse, or majority-noise assignments are discarded, regardless of their validity or stability scores. Rule 2 (cross-method consensus): among the non-degenerate solutions, the partition reproduced by the largest number of independent methods is identified as the candidate typology. Rule 3 (stability): the candidate partition must remain stable under data perturbation, and its stability must not be significantly exceeded by that of any non-degenerate alternative on the shared resamples. Rule 4 (representative method): among the methods producing the candidate partition, the representative is the one with the highest perturbation stability and the most interpretable structure. Internal validity indices inform but do not override this rule, and its application to the present data is reported step by step in Section 3.8. Stability itself was assessed as a data-perturbation sensitivity analysis rather than as an estimate of sampling variability, because the 38 provinces constitute the complete population of interest rather than a sample from a larger population; this perturbation view of stability follows the stability-based validation literature [13],

[24], [25], [26]. Two complementary perturbation schemes were applied, each with B = 50 replicates and full refitting of every method: the nonparametric bootstrap, which draws n = 38 provinces with replacement and is the standard perturbation device in this literature, although duplication and omission of observations can mechanically fragment very small clusters; and 90% subsampling, which draws m = 34 provinces without replacement [25] and perturbs the composition of the data while preserving the multiplicity of every retained observation. Because all methods are refitted on the same replicates, between-method stability comparisons are paired and were tested with two-sided Wilcoxon signed-rank tests. Cluster agreement between each refitted partition and its reference partition was measured using the ARI and NMI. ARI compares two partitions while correcting for chance agreement [27], while NMI measures the shared information between two clustering solutions [20].

Between-cluster differences on the original indicators were subsequently examined using permutation-based one-way tests, in which the observed F statistic was compared with its distribution under 10,000 random permutations of the cluster labels [28]. Welch's ANOVA [29] and Kruskal–Wallis tests were reported as supplementary evidence, and eta-squared and epsilon-squared were used as effect size measures. Because twelve indicators were tested simultaneously, Bonferroni correction was applied with an adjusted significance threshold of $\alpha_{adj} = 0.05/12 \approx 0.0042$; equivalently, Bonferroni-adjusted p-values were evaluated against the conventional 0.05 level. Since the smallest cluster contains only two provinces, parametric distributional assumptions are not tenable, and all these tests are interpreted as descriptive post-clustering diagnostics rather than confirmatory inference.

2.6 Implementation Details

All analyses were implemented in Python 3.13 using scikit-learn 1.8.0, SciPy 1.17.1, NumPy 2.4.4, pandas 3.0.2, hdbscan 0.8.42, fuzzy-c-means 0.0.6, and factor-analyzer 0.5.1, with a fixed random seed (42) for every stochastic procedure. K-Means used 50 random initializations. The hierarchical methods used Euclidean distance on the retained component scores. Fuzzy C-Means used fuzzifier m = 2. DBSCAN used min_samples = 4 with eps = 5.09 selected automatically at the knee of the sorted 4-nearest-neighbor distance curve; HDBSCAN used min_cluster_size = 3 and min_samples = 2. The Gaussian Mixture Model used full covariance matrices with 10 initializations. The ensemble procedure followed Algorithm 1 in Section 2.4. Bootstrap stability analysis drew B = 50 resamples with replacement; each method was refitted on every resample and compared, over the resampled observations (excluding noise assignments), with the reference partition obtained on the full data. The subsampling analysis drew B = 50 subsamples of 34 provinces (90%) without replacement under the same refitting and comparison protocol, and paired per-replicate stability comparisons used two-sided Wilcoxon signed-rank tests. Parallel analysis used 1,000 simulated standard-normal datasets of the same dimension as the observed data (n = 38, p = 12).

3. RESULTS AND DISCUSSION

3.1. Descriptive Statistics

Before conducting a more in-depth analysis, descriptive statistics were calculated to examine the characteristics of the data, as presented in Table 2. The results reveal substantial differences across variables in terms of dispersion and distributional patterns. X1 and X4 exhibit the highest variability, with coefficients of variation of 334.2% and 286.5%, respectively, indicating considerable heterogeneity among the observations. Relatively high variability is also observed in X7 and X6, suggesting that these variables are more dispersed compared with the others.

By contrast, X9, X10, and X3 show relatively low coefficients of variation, implying more stable distributions around their respective means. The skewness values further indicate that several variables deviate from a symmetric distribution. In particular, X1, X4, and X6 are positively skewed, reflecting a limited number of observations with substantially higher values. In contrast,

X5 and X11 show negative skewness, suggesting that their distributions are more concentrated toward higher values.

Table 2. Descriptive Statistics

Variables	Mean	SD	Median	Min	Max	CV (%)	Skewness
X1	51.06	170.64	0.70	0.16	766.64	334.2	3.44
X2	11.15	6.75	10.13	4.00	32.97	60.5	1.56
X3	51.38	4.81	51.49	39.66	66.97	9.4	0.37
X4	2.99	8.57	0.28	0.00	43.92	286.5	3.86
X5	8.61	1.35	8.72	3.40	11.19	15.7	-1.63
X6	13.00	11.82	10.99	0.04	69.36	90.9	3.07
X7	4.86	4.91	2.88	0.47	21.26	100.9	1.83
X8	20.80	5.83	20.95	7.00	36.60	28.0	0.53
X9	70.51	2.61	70.69	64.80	75.22	3.7	-0.15
X10	2004.00	148.19	2014.00	1696.00	2419.00	7.4	0.33
X11	59.93	6.69	60.58	39.89	74.45	11.2	-0.84
X12	45.22	8.46	45.77	26.92	61.22	18.7	-0.29

Figure 2 presents the boxplots of the twelve variables used in this study. Overall, the plots indicate that most variables are not symmetrically distributed, with several variables showing a concentration of observations at relatively lower values and a small number of observations extending toward higher values. This pattern is particularly evident in X1, X4, and X6, where the boxplots show long upper tails and extreme upper outliers. Such distributions suggest substantial disparities across provinces, where only a few observations record markedly higher values than the majority.

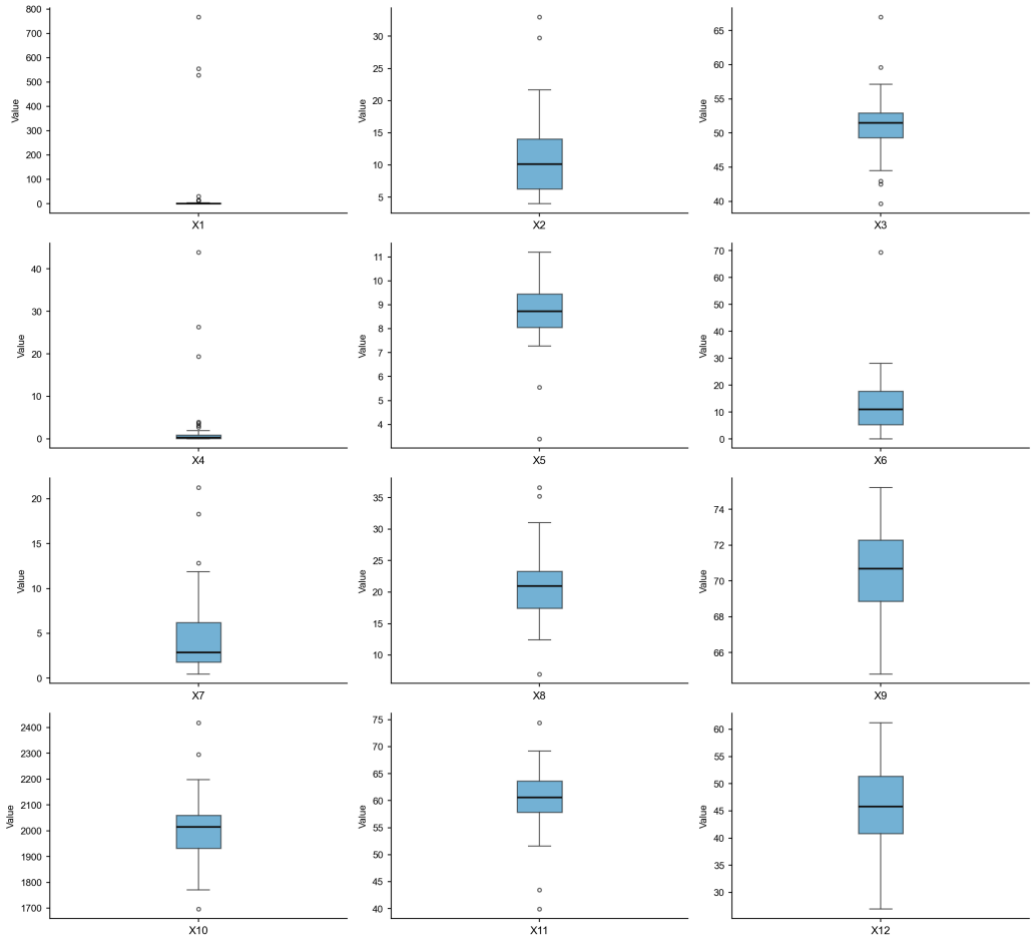


Figure 2. Boxplot distributions of the 12 food security indicators

A similar tendency, although less pronounced, can also be observed in X2 and X7, which show wider upper ranges and several high-value outliers. In contrast, variables such as X3, X8, X9, X10, X11, and X12 appear to have more compact interquartile ranges, indicating relatively lower dispersion among observations. However, outliers in several of these variables still suggest that certain provinces differ considerably from the general distribution. These findings are consistent with the descriptive statistics, particularly the high coefficients of variation and skewness observed in several variables. Furthermore, the provinces with the most extreme values on the vulnerability-related indicators, particularly X1, X4, and X6, are concentrated in eastern Indonesia, most notably in the Papua region, whereas provinces on Java Island generally record the most favorable values on protective indicators such as energy and protein consumption. This contrast indicates that food security-related conditions remain unevenly distributed across Indonesia.

3.2. Correlation Structure

Figure 3 presents the Pearson correlation heatmap among the twelve food security indicators. Overall, the correlation structure shows that several indicators are moderately to strongly associated, suggesting that food security conditions across provinces are shaped by interrelated dimensions rather than by isolated factors. Positive correlations are mainly observed among X2, X3, X4, X6, X7, and X8. For instance, X2 is strongly correlated with X4 and X8, while X6 also shows a strong positive association with X7. These relationships indicate that provinces with higher values in one indicator tend to show higher values in related indicators.

In contrast, several variables exhibit notable negative correlations. X5 is negatively correlated with X2, X3, X4, and X6, while X9 and X11 also show negative associations with several variables, particularly X2, X4, X6, X7, and X8. This pattern suggests that improvements in certain indicators may be accompanied by lower values in others, reflecting the multidimensional and potentially uneven nature of food security performance across regions. By contrast, X1 shows relatively weak correlations with most variables, indicating that it may capture a different aspect of food security compared with the other indicators.

The heatmap also indicates that no correlation coefficient exceeds the threshold for very high correlation, suggesting that the variables do not show severe redundancy. Nevertheless, several moderate to strong correlations highlight the need to consider interdependence among indicators in the subsequent analysis. Therefore, further analytical procedures are required to identify regional patterns more comprehensively and to classify Indonesian provinces based on the underlying characteristics of their food security profiles.

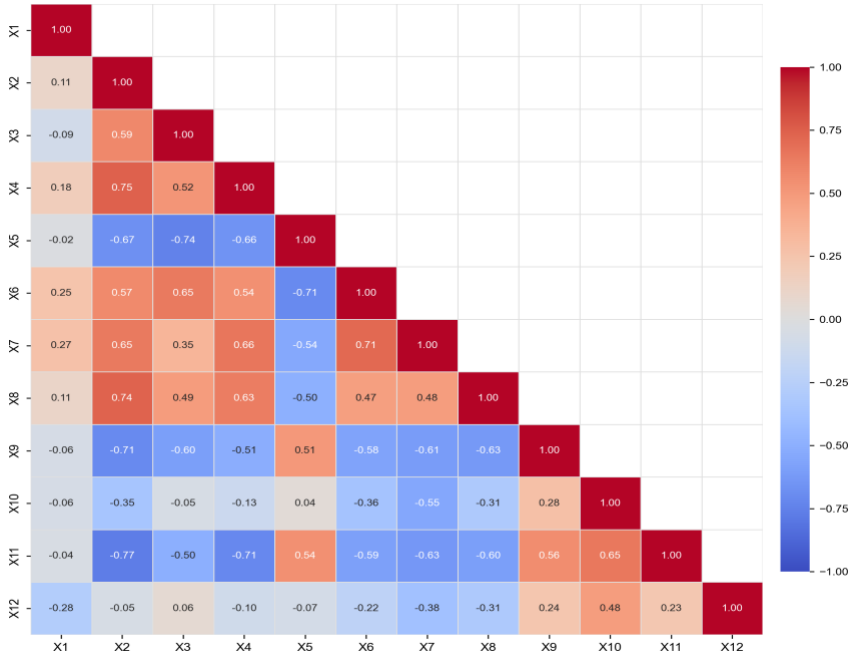


Figure 3. Pearson correlation heatmap of food security indicators

3.3. Principal Component Analysis (PCA)

Before cluster analysis, PCA was conducted to reduce the dimensionality of the twelve food security indicators. The adequacy of the data for PCA was evaluated using the KMO measure and Bartlett's test of sphericity. As shown in Table 3, Bartlett's test was statistically significant, $\chi^2 = 343.55$, df = 66, $p < 0.001$, indicating that the correlation matrix was not an identity matrix and that sufficient correlations existed among the variables. The overall KMO value was 0.673, which is considered acceptable for exploratory analysis. The individual KMO values for X1 (0.448) and X10 (0.377) were below the commonly recommended threshold of 0.50. Both variables were retained because they represent key dimensions of the food security framework, and the consequences of this decision were tested explicitly: a sensitivity analysis excluding X1 and X10, reported in Section 3.9, increased the overall KMO to 0.811 and produced a highly similar partition (ARI = 0.80) with the same two-province high-vulnerability cluster.

Table 3. KMO and Bartlett's test of sphericity Results

Test	Result
KMO	0.673
Bartlett's test of sphericity	$\chi^2 = 343.55$
	df = 66
	$p < 0.001$

Based on the eigenvalue criterion, three principal components were retained. As reported in Table 4, PC1 explains 50.68% of the total variance, while PC2 and PC3 account for 15.01% and 8.96%, respectively. Together, these three components explain 74.65% of the total variance, suggesting that the retained components capture a substantial proportion of the information contained in the original variables.

Table 4. Observed Eigenvalues, Parallel-Analysis Simulated Eigenvalues, and Variance Explained

Component	Eigenvalue	Simulated Mean	Simulated 95th Pct.	Variance (%)	Cumulative (%)
PC1	6.25	2.06	2.34	50.68	50.68
PC2	1.85	1.74	1.94	15.01	65.68
PC3	1.10	1.50	1.66	8.96	74.65
PC4	0.76	1.31	1.44	6.14	80.78
PC5	0.72	1.14	1.26	5.85	86.63
PC6	0.47	0.99	1.09	3.80	90.43
PC7	0.43	0.85	0.95	3.52	93.95
PC8	0.28	0.72	0.82	2.29	96.24
PC9	0.19	0.60	0.69	1.55	97.79
PC10	0.14	0.48	0.58	1.16	98.95
PC11	0.10	0.37	0.47	0.81	99.76
PC12	0.03	0.26	0.36	0.24	100.00

Figure 4 compares the observed eigenvalues with the eigenvalues obtained from Horn's parallel analysis based on 1,000 random standard-normal datasets of the same dimension (n = 38, p = 12) [16]; the exact simulated values are reported in Table 4. Under the mean criterion, only the first two observed eigenvalues (6.25 and 1.85) exceed their simulated counterparts (2.06 and 1.74), while the third observed eigenvalue (1.10) falls below its simulated mean (1.50); under the more conservative 95th percentile criterion, only PC1 is clearly supported. Parallel analysis therefore does not support the retention of PC3. PC3 was nevertheless retained based on the Kaiser criterion (eigenvalue = $1.10 \geq 1$), its contribution to cumulative explained variance (74.65% with three components versus 65.68% with two), and its substantive interpretability. To ensure that this choice does not drive the final typology, a sensitivity analysis reported in Section 3.9 repeated the clustering on two retained components: the two-province high-vulnerability cluster was preserved, although the boundary between the relatively food-secure and moderately vulnerable groups shifted (ARI = 0.24 against the three-component solution). The number of retained components therefore affects the allocation of borderline provinces but not the identification of the most vulnerable cluster.

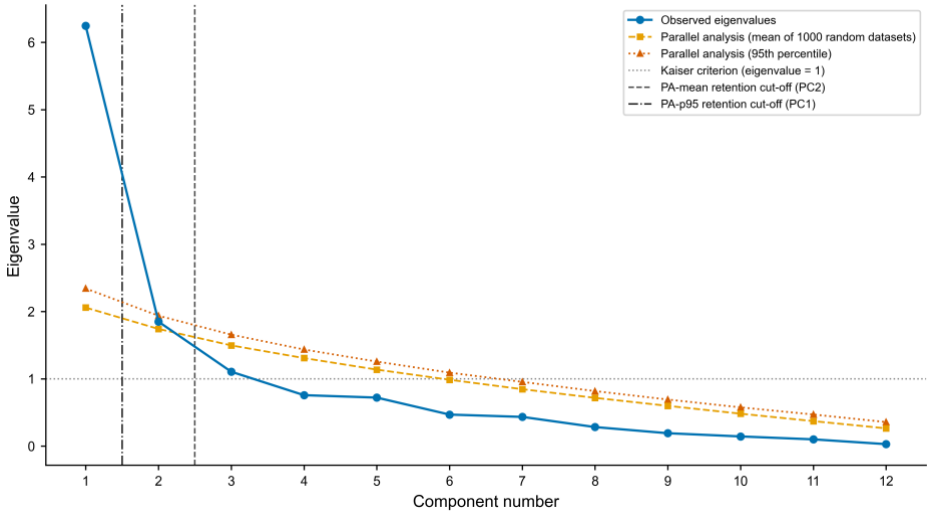


Figure 4. Scree plot of observed eigenvalues compared with parallel-analysis simulated eigenvalues (mean and 95th percentile of 1,000 random datasets)

3.4. Optimal Number of Clusters

The number of clusters was determined by combining the elbow method with three additional internal criteria computed for k = 2 to 10 on the retained component scores: the average Silhouette width, the Calinski–Harabasz index [30], and the gap statistic with 50 uniform reference datasets [31]. Figure 5 presents the four criteria. As is common in small samples with one strongly separated subgroup, the criteria do not converge on a single value of k. The WCSS curve shows its sharpest inflection at k = 3, where the reduction in WCSS is largest relative to subsequent steps (189.7 at k = 2, 123.7 at k = 3, and 82.2 at k = 4). The average Silhouette is maximized at k = 2 (0.617), but this two-cluster solution merely separates Papua Pegunungan and Papua Selatan from the remaining 36 provinces and provides no differentiation among the other provinces; the Silhouette values for k = 3 and k = 4 are 0.372 and 0.419, respectively. The Calinski–Harabasz index increases from k = 2 (28.6) to k = 3 (30.7) and continues to rise gradually without a clear optimum in the examined range, while the gap statistic first satisfies the standard-error rule at k = 5, a tendency toward larger k that is difficult to interpret with only 38 observations.

Because the statistical criteria disagree, the final choice balanced this evidence against substantive interpretability. No single internal index designates k = 3 as optimal, which is expected when one subgroup is strongly separated while the remaining provinces are comparatively similar: distance-based indices are then drawn either toward k = 2, which merely isolates that subgroup without differentiating the rest, or toward larger k. The value k = 3 was therefore justified not by the optimality of any single index but by the convergence of several considerations. It corresponds to the sharpest diminishing return in WCSS (the elbow) and improves the Calinski–Harabasz index relative to k = 2; it avoids the trivial two-group split that maximizes the Silhouette; and it yields a three-tier typology consistent with the priority classification used in Indonesia's Food Security and Vulnerability Atlas [5]. The criteria that favor larger k, in particular the gap statistic (k = 5), indicate that k = 3 is a conservative choice that does not over-partition the 38 provinces rather than a fragmentation of the data. Crucially, the core high-vulnerability structure does not depend on this choice: the two-province cluster recurs across neighboring values of k, across clustering methods (Section 3.5), and across preprocessing schemes (Section 3.9), so the selection of k = 3 governs the granularity of the typology rather than its central finding. The subjectivity inherent in this choice is acknowledged, and its consequences are examined through the multi-method comparison in Section 3.5 and the robustness analysis in Section 3.9.

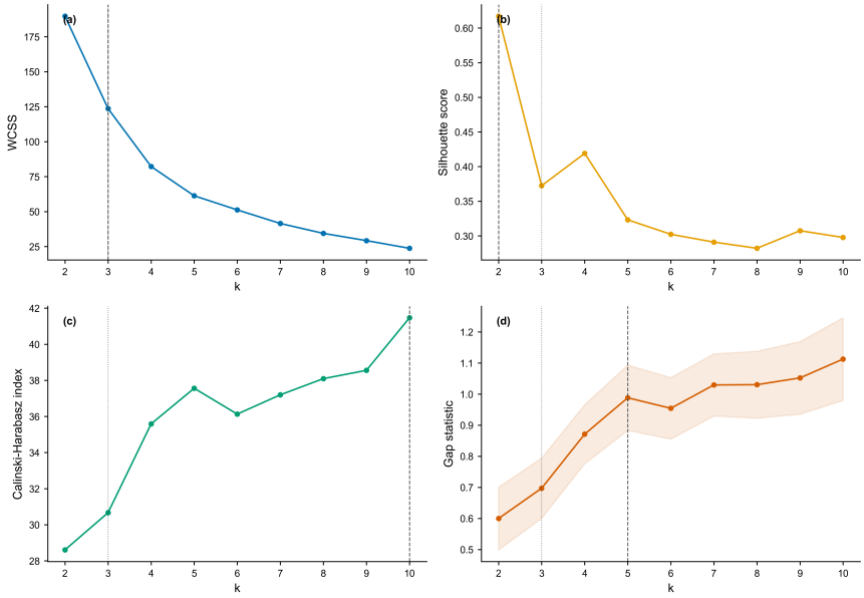


Figure 5. Internal criteria for selecting the number of clusters: (a) WCSS (elbow method), (b) average Silhouette width, (c) Calinski–Harabasz index, (d) gap statistic (shaded band: ±1 standard error)

3.5. Comparative Evaluation of Clustering Methods

Table 5 presents the comparative evaluation of ten clustering approaches applied to the three retained principal components. The methods include partitional, hierarchical, fuzzy, mixture-model, density-based, and consensus clustering approaches. For methods requiring a predefined number of clusters, the analysis was conducted using $k = 3$, following the result of the elbow method. In contrast, DBSCAN and HDBSCAN were treated as density-based methods with emergent cluster structures, as they do not require a fixed number of clusters.

Table 5. Comparative Evaluation of Multi-Method Clustering

Method	Type	k / emergent	n_clusters	n_noise	Silhouette	DBI	Dunn Index
K-Means	Partitional	3	3	0	0.372	0.771	0.100
Single Linkage	Hierarchical	3	3	0	0.577	0.232	0.520
Complete Linkage	Hierarchical	3	3	0	0.401	0.725	0.123
Average Linkage	Hierarchical	3	3	0	0.522	0.559	0.383
Ward Linkage	Hierarchical	3	3	0	0.401	0.725	0.123
Fuzzy C-Means	Fuzzy	3	3	0	0.346	0.935	0.053
DBSCAN	Density-based	emergent	1	1	-	-	-
HDBSCAN	Density-based	emergent	2	31	0.476	0.543	0.320
GMM	Mixture model	3	3	0	0.452	1.091	0.163
Ensemble Cluster	Consensus	3	3	0	0.401	0.725	0.123

The results show that each clustering approach produced different levels of performance. K-Means, as a partitional method, generated a moderate clustering structure, with a Silhouette coefficient of 0.372, a Davies–Bouldin Index of 0.771, and a Dunn Index of 0.100. These values indicate that K-Means formed three clusters, although the separation and compactness of the clusters were not as strong as those produced by some hierarchical methods.

Among the hierarchical methods, Single Linkage and Average Linkage produced the highest raw internal validity scores. Single Linkage achieved the highest Silhouette coefficient of 0.577, the lowest Davies–Bouldin Index of 0.232, and the highest Dunn Index of 0.520. Average Linkage also showed relatively strong performance, with a Silhouette coefficient of 0.522 and a Dunn Index of 0.383. However, these favorable values should be interpreted with caution because both methods generated highly imbalanced partitions. Single Linkage produced a 36–1–1 cluster structure, while Average Linkage resulted in a 34–2–2 structure. Such patterns indicate a chaining

effect, in which observations are gradually attached to one dominant cluster rather than forming substantively distinct groups [32]. Therefore, despite their high validity scores, Single Linkage and Average Linkage were not considered appropriate for the final cluster solution.

Fuzzy C-Means showed the weakest performance among the fixed-k methods, with the lowest Silhouette coefficient of 0.346, the highest Davies–Bouldin Index among fixed-k methods at 0.935, and the lowest Dunn Index of 0.053. This indicates weaker separation and lower compactness compared with the other non-density-based methods. Meanwhile, the Gaussian Mixture Model produced a moderate Silhouette coefficient of 0.452, but its higher Davies–Bouldin Index of 1.091 suggests less favorable cluster separation.

The density-based methods showed limited suitability for this dataset. DBSCAN produced only one dominant cluster and therefore did not provide a meaningful three-group partition for comparative interpretation. HDBSCAN generated two clusters but classified 31 provinces, or 81.6% of the observations, as noise. This result suggests that the density structure of the data in the PCA-reduced space was not sufficiently stable for density-based clustering, particularly given the relatively small sample size of 38 provinces [33].

The Ensemble clustering result produced the same validity values as Complete Linkage and Ward Linkage, with a Silhouette coefficient of 0.401, a Davies–Bouldin Index of 0.725, and a Dunn Index of 0.123. This indicates that the consensus solution was identical or highly similar to the Complete and Ward partitions. The convergence among Complete Linkage, Ward Linkage, and Ensemble clustering provides additional support for the non-degenerate three-cluster solution.

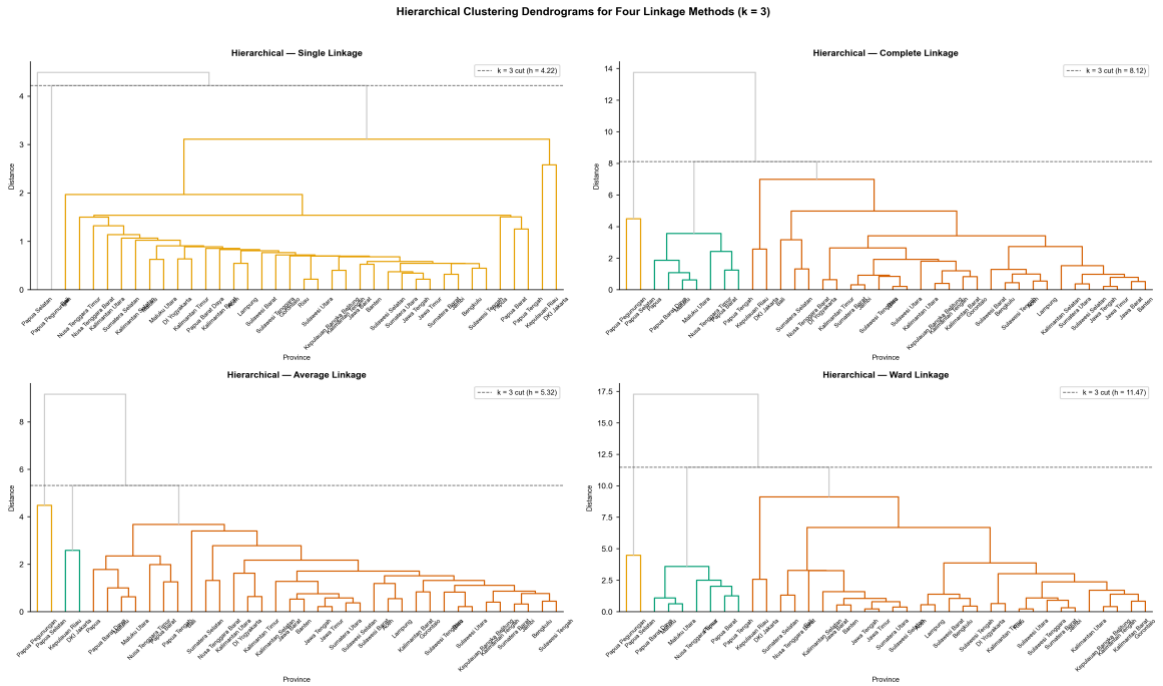


Figure 6. Dendrogram comparison of hierarchical clustering methods at $k = 3$

Figure 6 provides additional visual evidence regarding the effect of linkage choice on the resulting cluster structure. The Single Linkage dendrogram shows a gradual merging pattern with no clear separation among major branches. This pattern reflects the chaining effect and explains why Single Linkage produced a highly imbalanced 36–1–1 partition despite having the strongest internal validity scores.

The Average Linkage dendrogram shows a similar tendency, although the chaining pattern is less extreme than in Single Linkage. The resulting 34–2–2 partition indicates that most provinces were still absorbed into one dominant group, leaving only a very small number of provinces in the remaining clusters. Thus, the high Silhouette value of Average Linkage does not necessarily indicate a substantively meaningful provincial typology.

In contrast, the Complete Linkage dendrogram shows clearer separation around the $k = 3$ cut. The larger vertical gaps indicate that the clusters are more distinct and interpretable than those produced by Single and Average Linkage. Ward Linkage also shows a clearer hierarchical structure, with three groups separated at higher merge distances. This pattern supports the use of Ward Linkage as a more balanced solution because it forms clusters by minimizing within-cluster variation.

3.6. Cluster Membership by Method

A comparison of province-level assignments across the ten clustering methods reveals substantial agreement on the core vulnerability structure, alongside meaningful variation in how borderline provinces are allocated. Table 6 presents the cluster size distribution for each method using the original cluster labels, with noise assignments from the density-based methods reported in a separate column rather than as a cluster. Four methods produced degenerate partitions: Single Linkage and Average Linkage exhibited chaining artifacts, yielding imbalanced structures of 36–1–1 and 34–2–2, respectively; DBSCAN merged 37 of 38 provinces into a single cluster; and HDBSCAN classified 31 provinces (81.6%) as noise.

Table 6. Cluster Membership for All Clustering Methods

Method	Cluster 1	Cluster 2	Cluster 3	Noise
K-Means	9 (23.7%)	27 (71.1%)	2 (5.3%)	0
Single Linkage	36 (94.7%)	1 (2.6%)	1 (2.6%)	0
Complete Linkage	29 (76.3%)	2 (5.3%)	7 (18.4%)	0
Average Linkage	2 (5.3%)	34 (89.5%)	2 (5.3%)	0
Ward Linkage	29 (76.3%)	2 (5.3%)	7 (18.4%)	0
Fuzzy C-Means	23 (60.5%)	13 (34.2%)	2 (5.3%)	0
DBSCAN	37 (97.4%)	-	-	1 (2.6%)
HDBSCAN	4 (10.5%)	3 (7.9%)	-	31 (81.6%)
GMM	33 (86.8%)	3 (7.9%)	2 (5.3%)	0
Ensemble Cluster	29 (76.3%)	7 (18.4%)	2 (5.3%)	0

Among the six non-degenerate methods, Complete Linkage, Ward Linkage, and Ensemble clustering produced label-equivalent 29–7–2 partitions, although the numeric cluster labels differed across methods. K-Means yielded a 9-27-2 split, Fuzzy C-Means the most balanced partition at 23-13-2, and GMM the most concentrated at 33-3-2. The intermediate group thus ranges from n = 3 (GMM) to n = 13 (Fuzzy C-Means), indicating that several provinces lie close to the multivariate decision boundary. The smallest cluster, by contrast, is invariant across all six non-degenerate fixed-k methods: each of them isolates exactly the same two provinces, Papua Pegunungan and Papua Selatan. This consensus does not extend to the density-based methods, which do not produce an equivalent three-cluster structure: DBSCAN absorbed 37 provinces into a single cluster and HDBSCAN labeled most provinces as noise. Within the scope of the non-degenerate partitions, however, the cross-method agreement provides strong evidence of a structurally distinct high-vulnerability pair that is not an artifact of a particular clustering criterion; its robustness to outlier-related concerns is examined further in Section 3.9.

The degenerate methods reinforce the necessity of multi-method validation. Single Linkage and Average Linkage, despite achieving the highest internal validity scores in Table 5, produce substantively meaningless partitions due to the chaining effect [32]. DBSCAN and HDBSCAN confirm that density-based approaches require larger sample sizes to yield interpretable results. These cases demonstrate that high validity indices alone do not guarantee meaningful cluster structures.

3.7. Clustering Stability Analysis

To complement the internal validity evaluation, a bootstrap stability analysis was conducted using 50 resampling iterations as a robustness check. The stability of each clustering

method was assessed using the Adjusted Rand Index (ARI) and Normalized Mutual Information (NMI), where higher values indicate greater consistency of cluster assignments across bootstrap samples. Stability-based validation is widely used in unsupervised clustering because it evaluates the extent to which cluster structures are preserved under data perturbation or resampling [13], [24]. This analysis was not intended to replace the clustering evaluation presented in the previous section, but rather to examine whether the resulting partitions were robust to sampling variation.

Table 7. Stability Analysis Across Clustering Methods: Bootstrap Resampling (columns 2–5) and 90% Subsampling Without Replacement (columns 6–7)

Method	Mean of ARI	Standard Deviation of ARI	Mean of NMI	Standard Deviation of NMI	Subsampling ARI (mean ± SD)	Subsampling NMI (mean ± SD)
K-Means	0.612	0.318	0.678	0.232	0.900 ± 0.108	0.881 ± 0.120
Single Linkage	0.686	0.296	0.710	0.280	0.868 ± 0.268	0.878 ± 0.253
Complete Linkage	0.526	0.320	0.616	0.235	0.738 ± 0.322	0.788 ± 0.238
Average Linkage	0.885	0.247	0.901	0.200	0.968 ± 0.141	0.975 ± 0.107
Ward Linkage	0.527	0.362	0.632	0.251	0.804 ± 0.254	0.828 ± 0.191
Fuzzy C-Means	0.549	0.275	0.612	0.201	0.685 ± 0.226	0.721 ± 0.163
DBSCAN	0.000	0.000	0.000	0.000	0.000 ± 0.000	0.000 ± 0.000
HDBSCAN	0.439	0.436	0.498	0.424	0.423 ± 0.484	0.431 ± 0.485
GMM	0.540	0.277	0.556	0.238	0.801 ± 0.209	0.778 ± 0.198
Ensemble Cluster	0.498	0.315	0.605	0.224	0.793 ± 0.204	0.811 ± 0.164

As shown in Table 7, Average Linkage and Single Linkage produced the highest stability scores, with ARI values of 0.885 and 0.686, respectively. However, these results should be interpreted in relation to the findings in Section 3.5. Both methods generated degenerate and highly imbalanced cluster partitions, namely 34–2–2 for Average Linkage and 36–1–1 for Single Linkage. Therefore, although their bootstrap stability scores were high, the resulting clusters were not considered substantively meaningful for provincial typology. In this context, high stability reflects the repeated formation of the same imbalanced structure rather than the robustness of an interpretable clustering solution. Recent studies also emphasize that cluster stability should be interpreted together with cluster structure, validity indices, and substantive interpretability rather than used as a stand-alone criterion [12], [13].

Among the non-degenerate methods, K-Means showed the highest stability, with an ARI of 0.612 ± 0.318 and an NMI of 0.678 ± 0.232. This indicates that the K-Means solution was relatively consistent under bootstrap resampling. However, this result does not contradict the selection of the non-degenerate hierarchical and ensemble-based solution. The membership comparison shows that K-Means produced a partition that was strongly aligned with the Complete Linkage, Ward Linkage, and Ensemble solutions, with an ARI of 0.802 and an NMI of 0.775. The main difference lies in a small number of borderline provinces, particularly Gorontalo and Kalimantan Barat, suggesting that the overall three-cluster structure remains broadly consistent across alternative non-degenerate methods.

Complete Linkage, Ward Linkage, and Ensemble clustering produced identical or label-equivalent partitions, with a 29–7–2 cluster structure. This convergence provides additional support for the stability of the selected three-cluster solution. Moreover, Papua Pegunungan and Papua Selatan were consistently separated into a distinct small cluster across several non-degenerate methods, indicating that these provinces have a markedly different food security profile compared with the rest of the sample. The consistency across hierarchical and ensemble-based solutions further supports the use of multi-method evidence in evaluating clustering robustness [18].

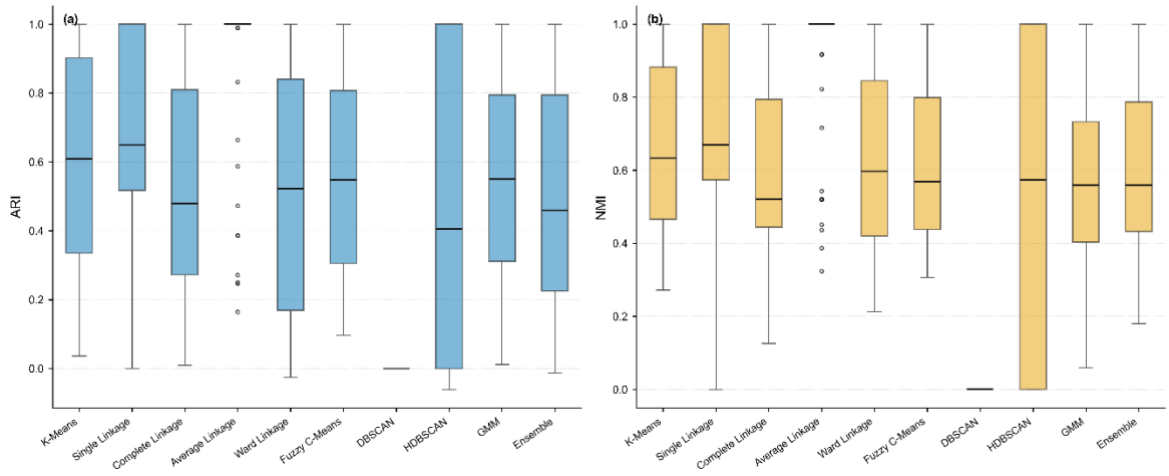


Figure 7. Boxplots of (a) Adjusted Rand Index (ARI) and (b) Normalized Mutual Information (NMI) across B = 50 bootstrap replications for all clustering methods.

The stability boxplots in Figure 7 illustrate the distribution of ARI and NMI scores across the 50 bootstrap replications. Average Linkage and Single Linkage show relatively high stability values; however, as discussed above, these results mainly reflect the repeated formation of imbalanced partitions rather than substantively meaningful cluster structures. Among the non-degenerate methods, K-Means shows relatively strong stability, as indicated by its comparatively high ARI and NMI distributions. Nevertheless, the spread of its stability values indicates that the solution remains somewhat sensitive to changes in sample composition. Ward Linkage and Complete Linkage also display moderate stability, although with wider distributions, suggesting variation in cluster assignments across bootstrap samples.

The density-based methods again showed limited suitability for this dataset. DBSCAN recorded zero stability because it repeatedly produced a single-cluster structure, making it unsuitable for meaningful partitioning. HDBSCAN showed moderate average stability, with an ARI of 0.439 ± 0.436 and an NMI of 0.498 ± 0.424, but the large standard deviations indicate substantial instability across bootstrap samples. This finding is consistent with the earlier result that HDBSCAN classified a large proportion of provinces as noise. Since density-based clustering relies on clear density separation, its performance may be less stable when the sample size is small, and the cluster structure is not sufficiently separated [34].

Although K-Means attains the highest bootstrap stability among the non-degenerate methods, this comparison must account for the paired design: every method was refitted on the same 50 resamples. On the shared replicates, the mean ARI advantage of K-Means over Ward Linkage is 0.085 (K-Means more stable in 24 replicates, Ward in 18, with 8 ties), and the paired difference is not statistically significant (two-sided Wilcoxon signed-rank test, $p = 0.088$; paired t-test, $p = 0.056$); the corresponding NMI difference (0.046) is likewise non-significant ($p = 0.228$). The nominal stability advantage of K-Means therefore lies within resampling noise.

Resampling with replacement also stresses a two-member cluster mechanically: duplication and omission mean that Papua Pegunungan and Papua Selatan are jointly present in only 20 of the 50 bootstrap resamples. Under the complementary 90% subsampling scheme (Table 7, columns 6–7), stability is uniformly higher for every method, the ordering of the non-degenerate methods is preserved (K-Means 0.900, Ward Linkage 0.804, Ensemble 0.793, Complete Linkage 0.738), and in all 42 subsamples containing both provinces the refitted Ward and ensemble solutions reproduced the two-member high-vulnerability cluster exactly. The low bootstrap stability of the hierarchical and ensemble solutions thus partly reflects an artifact of duplication acting on a very small cluster rather than genuine fragility of the final typology, and the stability conclusions do not depend on the choice of resampling scheme.

3.8. Optimal Method and Cluster Profiling

The final selection follows the decision rule specified in Section 2.5, applied step by step. Rule 1 (non-degeneracy) eliminates Single Linkage and Average Linkage, whose chaining yields 36–1–1 and 34–2–2 partitions, and the density-based methods, which collapse into a single cluster (DBSCAN) or assign 31 provinces to noise (HDBSCAN); their high validity or stability scores are therefore not decision-relevant. Rule 2 (cross-method consensus) identifies the 29–7–2 partition as the candidate typology: it is produced identically (pairwise ARI = 1.000) by Complete Linkage, Ward Linkage, and the ensemble consensus, which itself aggregates all seven non-degenerate base partitions, whereas no other partition is reproduced by more than one method. Rule 3 (stability) is satisfied: under 90% subsampling the refitted Ward solution reproduced the two-member high-vulnerability cluster in all 42 subsamples containing both provinces, and the only method with nominally higher stability, K-Means, does not significantly exceed the Ward solution on the paired replicates (Section 3.7), while its own partition agrees closely with the candidate typology (ARI = 0.802), differing only in the allocation of borderline provinces between Clusters I and II. Rule 4 selects Ward Linkage as the representative of the consensus partition: among the three methods producing it, Ward attains the highest subsampling stability (0.804, versus 0.793 for the ensemble and 0.738 for Complete Linkage) and provides a directly interpretable dendrogram through its variance-minimizing criterion. This convergence is evidence about the partition rather than about the intrinsic superiority of Ward as an algorithm; the substantive object of selection is the 29–7–2 typology itself. Accordingly, the following provincial typology was developed based on the Ward Linkage cluster membership.

The cluster profiling was based on the mean profile of each food security indicator, as presented in Table 8. Since the variables have different directions of interpretation, the cluster labels should be understood as relative typologies rather than absolute classifications.

Table 8. Average of Each Cluster

Variables	Cluster I	Cluster II	Cluster III
X1	45.4	7.8	284.2
X2	8.4	16.6	31.4
X3	50.6	51.0	63.3
X4	0.3	5.2	35.1
X5	8.9	8.7	4.5
X6	10.3	15.3	43.9
X7	3.1	8.7	16.6
X8	18.8	24.6	35.9
X9	71.5	67.8	65.9
X10	2,052	1,809	1,991
X11	62.7	53.6	41.7
X12	47.0	38.6	43.1

Table 8 and Figure 8 show that the three clusters have distinct indicator profiles and spatial patterns. Cluster I is predominantly located in western and central Indonesia, whereas Cluster II and Cluster III are concentrated in eastern Indonesia. This pattern suggests that food security-related disparities remain spatially uneven, particularly between provinces with stronger infrastructure, market access, and basic services and those facing more persistent geographical and structural constraints.

Cluster I is labeled as the relatively food-secure cluster. This cluster consists of 29 provinces, representing 76.3% of the provinces across Indonesia. It includes all provinces in Java, Bali, and Sumatra, as well as most provinces in Kalimantan and Sulawesi. As shown in Table 8, Cluster I records the lowest poverty rate at 8.4%, the lowest percentage of households without electricity at 0.3%, and the lowest stunting prevalence at 18.8%. It also has the highest female schooling level at 8.9 years, the highest life expectancy at 71.5 years, the highest energy consumption at 2,052 kcal/cap/day, and the highest protein consumption at 62.7 g/cap/day. These

characteristics indicate that provinces in Cluster I generally perform better in terms of food affordability, utilization, and consumption.



Figure 8. Cluster Profiling Maps

Cluster II represents a moderate vulnerability profile and consists of seven provinces, or 18.4% of the provinces in Indonesia. This cluster includes Nusa Tenggara Timur, Maluku, Maluku Utara, Papua Barat, Papua, Papua Barat Daya, and Papua Tengah. Compared with Cluster I, Cluster II has a higher poverty rate of 16.6%, higher stunting prevalence of 24.6%, and lower life expectancy of 67.8 years. It also records the lowest energy consumption among the three clusters, at 1,809 kcal/cap/day, and lower protein consumption at 53.6 g/cap/day. In terms of basic services, 15.3% of households in this cluster lack access to clean water, compared with 10.3% in Cluster I. These results indicate that Cluster II faces overlapping challenges related to food affordability, utilization, and consumption.

Cluster III represents the high-vulnerability cluster and consists of two provinces, namely Papua Pegunungan and Papua Selatan, accounting for 5.3% of the total provinces. This cluster shows the most severe vulnerability profile across several indicators. As reported in Table 8, Cluster III has the highest poverty rate at 31.4%, the highest percentage of households without electricity at 35.1%, the highest percentage of households without access to clean water at 43.9%, and the highest stunting prevalence at 35.9%. It also records the lowest female schooling level at 4.5 years, the lowest life expectancy at 65.9 years, and the lowest protein consumption at 41.7 g/cap/day. Although its energy consumption reaches 1,991 kcal/cap/day, which is higher than that of Cluster II, this does not necessarily indicate better overall food security. The lower protein intake suggests that food consumption in Cluster III may be constrained by limited dietary diversity and quality.

The high vulnerability of Cluster III is also reflected in the food availability indicator. The mean value of X1 reaches 284.2, substantially higher than those of Cluster I and Cluster II. Since X1 represents the ratio of normative per-capita consumption to food production, a higher value indicates a greater imbalance between food demand and local production capacity. Within the indicator framework used in this study, Papua Pegunungan and Papua Selatan therefore exhibit the most distinct vulnerability profile among the 38 provinces.

3.9. Robustness of the High-Vulnerability Cluster

Because Cluster III contains only two provinces, and because X1, X4, and X6 are highly skewed with extreme values (Table 2, Figure 2), additional analyses were conducted to examine whether this cluster is an artifact of extreme observations. The Ward solution was re-estimated under alternative preprocessing schemes, component dimensionalities, and variable subsets, and each resulting partition was compared with the baseline using the ARI, with particular attention to

whether Papua Pegunungan and Papua Selatan remained isolated as a distinct cluster. Table 9 summarizes the results.

Table 9. Robustness of the Ward Partition Under Alternative Specifications

Scenario	Specification	Cluster sizes	ARI vs baseline	Papua pair isolated
Two components	z-score; 2 PCs	18–18–2	0.239	Yes
Log transformation	log(1+x) on X1, X4, X6; z-score; 3 PCs	29–7–2	1.000	Yes
Winsorization	5th/95th percentile clipping; z-score; 3 PCs	28–8–2	0.708	No
Robust scaling	median/IQR scaling; 3 PCs	35–2–1	0.061	No
Skewed variables removed	X1, X4, X6 dropped; z-score; 2 PCs	25–11–2	0.076	Yes
Low-KMO variables removed	X1, X10 dropped; z-score; 2 PCs	27–9–2	0.802	Yes
Leave-one-variable-out	12 replications; z-score	varies	0.10–0.90	Yes (12 of 12)

For reduced variable sets, components with eigenvalue ≥ 1 were retained. Baseline: Ward Linkage on three principal components, partition 29–7–2.

Three findings support the substantive reality of the high-vulnerability cluster. First, applying a log transformation, log(1 + x), to the three most skewed indicators before standardization, which compresses extreme values while preserving their order, produced a partition identical to the baseline (ARI = 1.000); the cluster therefore does not depend on the leverage of raw extreme values. Second, the cluster survived the complete removal of the three skewed indicators: clustering on the remaining nine indicators still isolated the Papua pair, although the boundary between Clusters I and II shifted considerably (ARI = 0.076). Third, in all twelve leave-one-variable-out replications the Papua pair remained isolated (overall ARI between 0.10 and 0.90), indicating that no single indicator drives the result. The pair also remained isolated when only two principal components were retained (Section 3.3) and when the two indicators with individual KMO below 0.50 were excluded (ARI = 0.80, overall KMO rising from 0.673 to 0.811). Two scenarios did dissolve the cluster. Under winsorization at the 5th and 95th percentiles, the Papua pair merged into the broader eastern cluster; with 38 observations, however, clipping at these percentiles replaces approximately the two most extreme values of every indicator, which are precisely the genuinely measured conditions (for example, 43.9% of households without access to clean water) that distinguish these provinces, rather than measurement artifacts. Robust median/IQR scaling also failed to reproduce the structure (ARI = 0.061), which is attributable to indicators such as X4 whose interquartile range is close to zero, so that median/IQR scaling arbitrarily inflates negligible differences among non-vulnerable provinces.

Cluster-level diagnostics nonetheless indicate that the separation of the pair, while consistent, is quantitatively marginal: the Silhouette values of Papua Pegunungan and Papua Selatan in the baseline solution are 0.24 and 0.50, and the distance between the two provinces in the component space (4.49) is slightly larger than the minimum distance from the pair to the nearest province outside the cluster (4.22). Cluster III should therefore be interpreted as the extreme end of the eastern vulnerability continuum rather than as a categorically separate regime: under transformations that compress genuine extreme values, its members rejoin the moderate-vulnerability eastern cluster. This interpretation is consistent with the cluster profiling in Section 3.8 and does not alter the policy implication that these two provinces warrant the most intensive interventions.

3.10. Between-Cluster Difference Diagnostics

To examine the statistical separability of the three clusters, between-cluster differences were assessed for the twelve food security indicators, as reported in Table 10. Because Cluster III contains only two observations, the assumptions underlying parametric one-way ANOVA, namely normality and reliable within-group variance estimation, are untenable, and the Kruskal–Wallis test is likewise unstable under such extreme size imbalance. Permutation-based one-way tests were therefore used as the primary evidence: for each indicator, the observed F statistic was compared with its distribution under 10,000 random permutations of the cluster labels [28], a procedure that requires no distributional assumptions. Welch's ANOVA [29] and Kruskal–Wallis tests are reported as supplementary evidence, together with eta-squared and epsilon-squared effect sizes.

Since the clusters were derived from the same indicators, all results should be read as post-clustering descriptive diagnostics rather than confirmatory hypothesis tests. Given the twelve simultaneous tests, Bonferroni correction was applied with an adjusted significance threshold of $0.05/12 \approx 0.0042$; equivalently, Bonferroni-adjusted p-values were evaluated against the conventional 0.05 level.

Table 10. Permutation, Welch, and Kruskal–Wallis Diagnostics for Between-Cluster Differences

Variables	F	Perm. p	Perm. p (Bonf.)	η^2	H	KW p (Bonf.)	ε^2
X1	2.248	0.133	1.000	0.114	7.488	0.284	0.202
X2	49.059	< 0.001	0.001	0.737	15.320	0.006	0.414
X3	9.454	0.001	0.017	0.351	5.620	0.722	0.152
X4	102.003	< 0.001	0.001	0.854	20.352	< 0.001	0.550
X5	20.478	< 0.001	0.008	0.539	5.800	0.660	0.157
X6	12.571	0.008	0.095	0.418	6.955	0.371	0.188
X7	18.690	0.002	0.018	0.516	11.408	0.040	0.308
X8	20.101	< 0.001	0.001	0.535	13.839	0.012	0.374
X9	16.503	< 0.001	0.001	0.485	16.562	0.003	0.448
X10	12.205	< 0.001	0.001	0.411	14.858	0.007	0.402
X11	42.453	< 0.001	0.001	0.708	20.042	< 0.001	0.542
X12	3.136	0.055	0.666	0.152	6.460	0.475	0.175

Table 10 shows that nine of the twelve indicators significantly differentiate the three clusters based on Bonferroni-adjusted permutation p-values: X2, X3, X4, X5, X7, X8, X9, X10, and X11. The strongest between-cluster differences are observed for X4, households without access to electricity ($\eta^2 = 0.854$), followed by X2, the poverty rate ($\eta^2 = 0.737$), and X11, protein consumption ($\eta^2 = 0.708$). These large effect sizes indicate that infrastructure access, income poverty, and dietary adequacy constitute the main dimensions separating the provincial clusters. Female schooling (X5), the population-to-health-worker ratio (X7), stunting prevalence (X8), life expectancy (X9), energy consumption (X10), and the food expenditure burden (X3) also differ significantly, with moderate to large effect sizes.

Three indicators do not reach statistical significance after Bonferroni correction under the permutation test: X1, the food availability ratio; X6, households without access to clean water; and X12, rice productivity. For X1, the non-significance reflects its extreme within-cluster variability; for X6, the large descriptive difference between cluster means (10.3% in Cluster I versus 43.9% in Cluster III) is supported by only two observations in the extreme group, so the permutation distribution remains wide; for X12, rice productivity alone does not follow the broader vulnerability gradient captured by the cluster structure. The Kruskal–Wallis results, reported as supplementary evidence, are broadly consistent, with X2, X4, X7, X8, X9, X10, and X11 remaining significant after correction.

A secondary comparison restricted to Cluster I ($n = 29$) and Cluster II ($n = 7$), using Mann–Whitney tests with Bonferroni correction, found significant differences for X2, X4, X8, X9, X10, and X11 but not for X3, X5, and X7. This indicates that the omnibus effects of X3, X5, and X7 are driven mainly by the extreme values of Cluster III rather than by the boundary between the relatively food-secure and moderately vulnerable groups. Overall, the diagnostics support the interpretation that the three clusters are meaningfully differentiated across key dimensions of food affordability, utilization, and consumption, while inference involving the two-province cluster remains descriptive.

Although the statistical tests in Table 10 identify which indicators significantly differentiate the clusters, they do not fully show the direction and relative pattern of differences across indicators. Therefore, a radar chart was used to complement the significance tests by visualizing the standardized mean profile of each cluster. This visualization helps illustrate how the clusters differ simultaneously across the twelve food security indicators and provides a clearer interpretation of the multidimensional structure underlying the provincial typology (Figure 9).

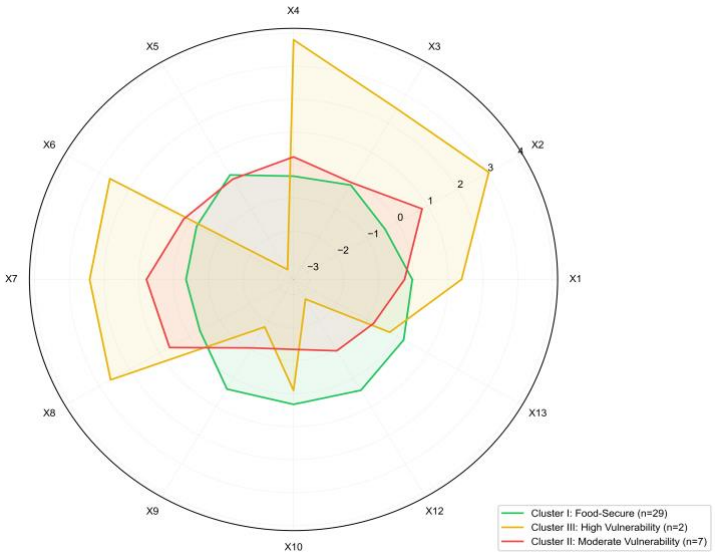


Figure 9. Radar chart of standardized food security profiles by cluster

Figure 9 presents the Z-standardized mean profiles of the three clusters across the twelve food security indicators. The radar chart should be interpreted by indicator rather than by overall polygon size, as the variables differ in direction. Higher values for X1–X4 and X6–X8 indicate greater vulnerability, whereas higher values for X5 and X9–X12 reflect more favorable conditions.

Cluster I is generally positioned close to the national average but shows a more favorable profile on protective indicators, including female schooling (X5), life expectancy (X9), energy consumption (X10), protein consumption (X11), and rice productivity (X12). It also records lower values on key vulnerability indicators, particularly poverty (X2), electricity deprivation (X4), clean water deprivation (X6), health-worker burden (X7), and stunting prevalence (X8). This pattern supports its classification as the relatively food-secure cluster.

Cluster II exhibits a moderate vulnerability profile, characterized by above-average poverty (X2), health-worker burden (X7), and stunting prevalence (X8), alongside below-average life expectancy (X9), energy consumption (X10), protein consumption (X11), and rice productivity (X12). These patterns indicate overlapping constraints in food affordability, utilization, and consumption, although less severe than those observed in Cluster III.

Cluster III displays the most distinct vulnerability profile. It records substantially above-average values for poverty (X2), food expenditure burden (X3), electricity deprivation (X4), clean water deprivation (X6), health-worker burden (X7), and stunting prevalence (X8), while female schooling (X5), life expectancy (X9), protein consumption (X11), and rice productivity (X12) fall below the national average. Although energy consumption (X10) is close to the national average and higher than in Cluster II, the low protein intake indicates that vulnerability in this cluster is driven not only by caloric availability, but also by limited dietary quality and broader structural deprivation.

4. CONCLUSION

This study compared multiple clustering algorithms to classify Indonesian provinces based on twelve food security indicators covering food availability, affordability, utilization, and consumption. PCA was applied prior to clustering, and three principal components were retained, explaining 74.65% of the total variance. The number of clusters was selected by combining the elbow criterion with the average Silhouette, Calinski–Harabasz, and gap statistics, together with substantive interpretability, leading to a three-cluster solution. Ward Linkage was selected as the best clustering method, as it produced the most interpretable and non-degenerate three-cluster solution. This selection was supported by Complete Linkage and Ensemble clustering, which yielded identical or highly similar partitions, while Single and Average Linkage were excluded due to chaining effects.

The final typology reveals clear spatial disparities in food security conditions. Cluster I represents a relatively food-secure group, mainly consisting of provinces in Java, Bali, Sumatra, and most of Kalimantan and Sulawesi. Cluster II reflects moderate vulnerability and includes several provinces in eastern Indonesia with higher poverty and stunting but lower food consumption indicators. Cluster III, consisting of Papua Pegunungan and Papua Selatan, shows the highest vulnerability profile, characterized by severe poverty, limited access to electricity and clean water, high stunting, low female schooling, low life expectancy, and low protein consumption. Permutation-based diagnostics support the interpretation that the typology captures meaningful differences, with electricity access, poverty, and protein consumption as the strongest differentiating factors. The two-province high-vulnerability cluster proved robust to log-transformation of skewed indicators, removal of individual variables, reduced component dimensionality, and exclusion of low-KMO variables, although its separation from the moderate-vulnerability cluster is quantitatively marginal. These findings show that food security disparities in Indonesia are multidimensional and spatially uneven, requiring targeted interventions in vulnerable provinces, particularly in eastern Indonesia.

This study has several limitations: it relies on aggregate provincial-level data, which may mask disparities within provinces; the internal criteria for selecting the number of clusters and components diverge, so the reported typology reflects a reasoned choice among defensible alternatives; and the smallest cluster contains only two provinces, which restricts formal inference. Future research should consider district-level data, temporal analysis, and spatial modeling to better capture the dynamics and drivers of food security vulnerability in Indonesia.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of the Institute for Research and Community Service, Universitas Terbuka, through the research grant under contract number B/361/UN31.LPPM/PT.01.03/2025.

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

DATA AVAILABILITY

The publicly available data used in this study were obtained from official statistics published by Statistics Indonesia (BPS), the National Food Agency (BAPANAS), and the Ministry of Health. The compiled and processed analytical dataset supporting the findings of this study is available from the corresponding author upon reasonable request.

AUTHOR CONTRIBUTIONS

First Author: Conceptualization; Data Curation; Methodology; Software; Formal Analysis; Visualization; Writing-Original Draft. **Second Author**: Conceptualization; Methodology; Validation; Supervision; Writing-Review & Editing. **Third Author**: Methodology; Validation; Supervision; Writing-Review & Editing.

REFERENCES

[1] FAO, IFAD, UNICEF, WFP, and WHO, *The State of Food Security and Nutrition in the World 2023*. FAO; IFAD; UNICEF; WFP; WHO, 2023. doi: 10.4060/cc3017en.

[2] C. B. Barrett, "Measuring food insecurity," *Science (1979).*, vol. 327, no. 5967, pp. 825–828, 2010, doi: 10.1126/science.1182768.

[3] A. D. Jones, F. M. Ngure, G. Pelto, and S. L. Young, "What are we assessing when we measure food security? A compendium and review of current metrics," *Advances in Nutrition*, vol. 4, no. 5, pp. 481–505, 2013, doi: 10.3945/an.113.004119.

[4] J. B. Upton, J. D. Cissé, and C. B. Barrett, "Food security as resilience: Reconciling definition and measurement," *Agricultural Economics*, vol. 47, no. S1, pp. 135–147, 2016, doi: 10.1111/agec.12305.

[5] Badan Pangan Nasional, "Peta Ketahanan dan Kerentanan Pangan (Food Security and Vulnerability Atlas/FSVA)." Accessed: Jul. 22, 2026. [Online]. Available: https://fsva.badanpangan.go.id/

[6] E. R. Amrullah, A. Ishida, A. Pullaila, and A. Rusyiana, "Who suffers from food insecurity in Indonesia?," *Int. J. Soc. Econ.*, vol. 46, no. 10, pp. 1186–1197, Oct. 2019, doi: 10.1108/IJSE-03-2019-0196.

[7] A. Akbar, R. Darma, I. M. Fahmid, and A. Irawan, "Determinants of Household Food Security during the COVID-19 Pandemic in Indonesia," *Sustainability*, vol. 15, no. 5, p. 4131, Feb. 2023, doi: 10.3390/su15054131.

[8] E. R. Amrullah, H. Tokuda, A. Rusyiana, and A. Ishida, "Effect of COVID-19 pandemic on food insecurity in Indonesian households," *Int. J. Soc. Econ.*, vol. 50, no. 12, pp. 1790–1803, 2023, doi: 10.1108/IJSE-03-2023-0186.

[9] R. Abdulah, "Mapping Food Security in Indonesia: Geographic Clusters and Regional Disparities," *Indonesian Journal of Geography*, vol. 57, no. 3, p. 440, Nov. 2025, doi: 10.22146/ijg.99419.

[10] R. S. Rahayu, Y. Purwaningsih, and A. Daerobi, "Mapping of provincial food security in Indonesia using based clustering model," *Jurnal Ekonomi Pembangunan: Kajian Masalah Ekonomi dan Pembangunan*, vol. 20, no. 1, pp. 69–79, 2019, doi: 10.23917/jep.v20i1.7096.

[11] V. Ratnasari and A. T. R. Dani, "Mapping the provincial food security conditions in Indonesia using cluster ensemble-based mixed data clustering-Robust clustering with links (CEBMDC-ROCK)," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 13, no. 2, pp. 611–617, 2023, doi: 10.18517/ijaseit.13.2.16457.

[12] M. Gagolewski, M. Bartoszuk, and A. Cena, "Are cluster validity measures (in)valid?," *Inf. Sci. (N. Y).*, vol. 581, pp. 620–636, 2021, doi: 10.1016/j.ins.2021.10.004.

[13] T. Liu, H. Yu, and R. H. Blair, "Stability estimation for unsupervised clustering: A review," *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 14, no. 6, p. e1575, 2022, doi: 10.1002/wics.1575.

[14] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, p. 20150202, Apr. 2016, doi: 10.1098/rsta.2015.0202.

[15] F. L. Gewers *et al.*, "Principal component analysis: A natural approach to data exploration," *ACM Comput. Surv.*, vol. 54, no. 4, May 2021, doi: 10.1145/3447755.

[16] J. L. Horn, "A rationale and test for the number of factors in factor analysis," *Psychometrika*, vol. 30, no. 2, pp. 179–185, 1965, doi: 10.1007/BF02289447.

[17] D. Xu and Y. Tian, "A Comprehensive Survey of Clustering Algorithms," *Annals of Data Science*, vol. 2, no. 2, pp. 165–193, Jun. 2015, doi: 10.1007/s40745-015-0040-1.

[18] A. A. Wani, "Comprehensive analysis of clustering algorithms: Exploring limitations and innovative solutions," *PeerJ Comput. Sci.*, vol. 10, p. e2286, 2024, doi: 10.7717/peerj-cs.2286.

[19] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and H. Jia, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Inf. Sci. (N. Y).*, vol. 622, pp. 178–210, 2023, doi: 10.1016/j.ins.2022.11.139.

[20] A. Strehl and J. Ghosh, "Cluster ensembles: A knowledge reuse framework for combining multiple partitions," *Journal of Machine Learning Research*, vol. 3, pp. 583–617, 2002, [Online]. Available: https://www.jmlr.org/papers/v3/strehl02a.html

[21] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, 1987, doi: 10.1016/0377-0427(87)90125-7.

[22] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, pp. 224–227, 1979, doi: 10.1109/TPAMI.1979.4766909.

[23] J. C. Dunn, "Well-separated clusters and optimal fuzzy partitions," *Journal of Cybernetics*, vol. 4, no. 1, pp. 95–104, 1974, doi: 10.1080/01969727408546059.

[24] H. Yu *et al.*, "Bootstrapping estimates of stability for clusters, observations and model selection," *Comput. Stat.*, vol. 34, no. 1, pp. 349–372, Mar. 2019, doi: 10.1007/s00180-018-0830-y.

[25] C. Hennig, "Cluster-wise assessment of cluster stability," *Comput. Stat. Data Anal.*, vol. 52, no. 1, pp. 258–271, Sep. 2007, doi: 10.1016/j.csda.2006.11.025.

[26] U. von Luxburg, "Clustering Stability: An Overview," *Foundations and Trends® in Machine Learning*, vol. 2, no. 3, pp. 235–274, Apr. 2010, doi: 10.1561/2200000008.

[27] L. Hubert and P. Arabie, "Comparing partitions," *J. Classif.*, vol. 2, no. 1, pp. 193–218, Dec. 1985, doi: 10.1007/BF01908075.

[28] M. D. Ernst, "Permutation methods: A basis for exact inference," *Statistical Science*, vol. 19, no. 4, pp. 676–685, 2004, doi: 10.1214/088342304000000396.

[29] B. L. Welch, "On the comparison of several mean values: An alternative approach," *Biometrika*, vol. 38, no. 3/4, pp. 330–336, 1951, doi: 10.2307/2332579.

[30] T. Calinski and J. Harabasz, "A dendrite method for cluster analysis," *Communications in Statistics*, vol. 3, no. 1, pp. 1–27, 1974, doi: 10.1080/03610927408827101.

[31] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic," *J. R. Stat. Soc. Series B Stat. Methodol.*, vol. 63, no. 2, pp. 411–423, 2001, doi: 10.1111/1467-9868.00293.

[32] F. Murtagh and P. Legendre, "Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion?," *J. Classif.*, vol. 31, no. 3, pp. 274–295, Oct. 2014, doi: 10.1007/s00357-014-9161-z.

[33] R. J. G. B. Campello, D. Moulavi, and J. Sander, "Density-Based Clustering Based on Hierarchical Density Estimates," in *Advances in Knowledge Discovery and Data Mining*, 2013, pp. 160–172. doi: 10.1007/978-3-642-37456-2_14.

[34] E. S. Dalmaijer, C. L. Nord, and D. E. Astle, "Statistical power for cluster analysis," *BMC Bioinformatics*, vol. 23, no. 1, p. 205, Dec. 2022, doi: 10.1186/s12859-022-04675-1.